

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ



НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ

САНКТ-ПЕТЕРБУРГСКОГО ГОСУДАРСТВЕННОГО
ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

Информатика. Телекоммуникации.
Управление

5(229) 2015

Издательство Политехнического университета
Санкт-Петербург
2015

НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ САНКТ-ПЕТЕРБУРГСКОГО
ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА
ИНФОРМАТИКА. ТЕЛЕКОММУНИКАЦИИ. УПРАВЛЕНИЕ

РЕДАКЦИОННЫЙ СОВЕТ ЖУРНАЛА

Председатель

Юсупов Р.М., чл.-кор. РАН;

Редакционный совет:

Абрамов С.М., чл.-кор. РАН;

Арсеньев Д.Г., д-р техн. наук, профессор;

Воеводин В.В., чл.-кор. РАН;

Заборовский В.С., д-р техн. наук, профессор;

Козлов В.Н., д-р техн. наук, профессор;

Фотиади А.Э., д-р физ.-мат. наук, профессор;

Черноруцкий И.Г., д-р техн. наук, профессор.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ ЖУРНАЛА

Главный редактор

Коротков А.С., д-р техн. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Редакционная коллегия:

Бабкин А.В., д-р экон. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Ицыксон В.М., канд. техн. наук, доцент, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Prof. Dr. Philippe Ferrari, Head of the RF and Millimeter-Wave Lab IMEP-LAHC Microelectronics, Electromagnetism and Photonic Institute, Grenoble Alpes University, France;

Карнов Ю.Г., д-р техн. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Клавдиев В.Е., канд. техн. наук, доцент, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Prof. Dr. Wolfgang Krautschneider, Head of Nanoelectronics Institute, Hamburg University of Technology, Germany;

Кучерявый Е.А., канд. техн. наук, профессор, Tampere University of Technology, Finland.

Dr. Fa-Long Luo, Chief Scientist, Element CXI, San Jose, USA;

Макаров С.Б., д-р техн. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Prof. Dr. Emil Novakov, IMEP-LAHC Microelectronics, Electromagnetism and Photonic Institute, Grenoble, France;

Устинов С.М., д-р техн. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Цикин И.А., д-р техн. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Шкодывев В.П., д-р техн. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Россия.

Журнал с 1995 года издается под научно-методическим руководством Российской академии наук. С 2008 года выпускается в составе сериального периодического издания «Научно-технические ведомости СПбГПУ» ISSN 1994-2354.

Журнал с 2002 года входит в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

Сведения о публикациях представлены в Реферативном журнале ВИНТИ РАН, в международной справочной системе «Ulrich's Periodical Directory».

Журнал зарегистрирован Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации ПИ № ФС77-51457 от 19.10.2012 г.

Подписной индекс **47517** в объединенном каталоге «Пресса России».

Журнал включен в базу данных «Российский индекс научного цитирования» (РИНЦ), размещенную на платформе Научной электронной библиотеки на сайте <http://www.elibrary.ru>

При перепечатке материалов ссылка на журнал обязательна.

Точка зрения редакции может не совпадать с мнением авторов статей.

Адрес редакции и издательства: Россия, 195251, Санкт-Петербург, ул. Политехническая, д. 29.

Тел. редакции (812) 552-62-16.

THE MINISTRY OF EDUCATION AND SCIENCE OF THE RUSSIAN FEDERATION



**ST. PETERSBURG STATE
POLYTECHNICAL UNIVERSITY
JOURNAL**

Computer Science.
Telecommunications and Control Systems

5(229) 2015

Polytechnical University Publishing House
Saint Petersburg
2015

ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL
COMPUTER SCIENCE. TELECOMMUNICATIONS AND CONTROL SYSTEMS

EDITORIAL COUNCIL

Head of the editorial council

Prof. Dr. *Rafael M. Yusupov* (corresponding member of the Russian Academy of Sciences)

Members:

Prof. Dr. *Sergey M. Abramov* (corresponding member of the Russian Academy of Sciences),

Prof. Dr. *Dmitry G. Arseniev*,

Prof. Dr. *Vladimir V. Voevodin* (corresponding member of the Russian Academy of Sciences),

Prof. Dr. *Vladimir S. Zaborovsky*,

Prof. Dr. *Vladimir N. Kozlov*,

Prof. Dr. *Alexandr E. Fotiadi*,

Prof. Dr. *Igor G. Chernorutsky*.

EDITORIAL BOARD

Editor-in-chief

Prof. Dr. *Alexander S. Korotkov*, Peter the Great St. Petersburg Polytechnic University, Russia;

Members:

Prof. Dr. *Alexandr V. Babkin*, Peter the Great St. Petersburg Polytechnic University, Russia;

Assoc. Prof. Dr. *Vladimir M. Itsykson*, Peter the Great St. Petersburg Polytechnic University, Russia;

Prof. Dr. *Philippe Ferrari*, Head of the RF and Millimeter-Wave Lab IMEP-LAHC Microelectronics, Electromagnetism and Photonic Institute, Grenoble Alpes University, France;

Prof. Dr. *Yuri G. Karpov*, Peter the Great St. Petersburg Polytechnic University, Russia;

Assoc. Prof. Dr. *Vladimir E. Klavdiev*, Peter the Great St. Petersburg Polytechnic University, Russia;

Prof. Dr. *Yevgeni Koucheryavy*, Tampere University of Technology, Finland.

Prof. Dr. *Wolfgang Krautschneider*, Head of Nanoelectronics Institute, Hamburg University of Technology, Germany;

Dr. *Fa-Long Luo*, Chief Scientist, Element CXI, San Jose, USA;

Prof. Dr. *Sergey B. Makarov*, Peter the Great St. Petersburg Polytechnic University, Russia;

Prof. Dr. *Emil Novakov*, IMEP-LAHC Microelectronics, Electromagnetism and Photonic Institute, Grenoble, France;

Prof. Dr. *Viacheslav P. Shkodyrev*, Peter the Great St. Petersburg Polytechnic University, Russia;

Prof. Dr. *Igor A. Tsikin*, Professor, Peter the Great St. Petersburg Polytechnic University, Russia;

Prof. Dr. *Sergey M. Ustinov*, Peter the Great St. Petersburg Polytechnic University, Russia.

The journal is published under scientific and methodical guidance of the Russian Academy of Sciences since 1995. The journal is published since 2008 as part of the periodical edition «Nauchno-tekhnicheskie vedomosti SPbGPU» (ISSN 1994-2354).

The journal is included in the List of Leading Peer-Reviewed Scientific Journals and other editions to publish major findings of PhD theses for the research degrees of Doctor of Sciences and Candidate of Sciences.

The publications are presented in the VINITI RAS Abstract Journal and Ulrich's Periodical Directory International Database.

The journal is registered with the Federal Service for Supervision in the Sphere of Telecom, Information Technologies and Mass Communications (ROSKOMNADZOR). Certificate ПИ № ФС77-51457 issued Oct. 19, 2012.

Subscription index **47517** in the «Press of Russia» Joint Catalogue.

The journal is on the Russian Science Citation Index (RSCI) database

© Scientific Electronic Library (<http://elibrary.ru/>).

No part of this publication may be reproduced without clear reference to the source.

The views of the authors can contradict the views of the Editorial Board.

The address: 195251 Polytekhnicheskaya Str. 29, St. Petersburg, Russia.

© Peter the Great St. Petersburg Polytechnic University, 2015

Содержание

Информационные технологии

Автюшенко А.Л., Иванов В.М. *Компьютерные технологии конвертации двухмерных объектов в трехмерную сцену* 7

Моделирование вычислительных, телекоммуникационных, управляющих и социально-экономических систем

Бабакина Н.А., Колесников М.П., Шкодырев В.П. *Представление окружающего пространства на основе векторных динамических структур в мобильных робототехнических системах* 19

Программное обеспечение вычислительных, телекоммуникационных и управляющих систем

Никулина Е.А. *Расчет функции рассеяния точки с учетом двулучепреломления и частичной когерентности освещения с использованием графических процессоров* 33

Аппаратное обеспечение вычислительных, телекоммуникационных и управляющих систем

Никитин А.Б., Хабитуева Е.И. *СВЧ-модель бескорпусного резистора* 39

Информационные, управляющие и измерительные системы

Христодуло О.И., Салимзянов И.Ф., Гареева Н.Р. *Разработка информационной системы размещения объектов техногенной опасности с использованием нечеткой логики* 47

Малыхина Г.Ф., Кислицына И.А. *Измерение параметров движения с использованием нейронных сетей* 59

Интеллектуальные системы и технологии

Соколов К.В., Тимофеев Д.А., Самочадин А.В. *Извлечение описаний бизнес-процессов из текстов на естественном языке* 69

Корелин В.Н., Блеканов И.С., Сергеев С.Л. *Применение модифицированного алгоритма LSH для кластеризации внешнего окружения веб-пространства университетов* 79

Хасина Е.М., Ломов А.А. *Сжатие аудиофайлов методом STLS-ESM* 88

Инфокоммуникационные технологии в образовании

Маслов М.Ю., Носницын С.М., Самочадин А.В., Логинов К.Е. *Архитектура средств управления мобильными устройствами для образовательных учреждений* 97

Contents

Information Technologies

Avtiushenko A.L., Ivanov V.M. <i>Computer Technologies for Converting Two-Dimensional Objects to the Three-Dimensional Scene</i>	7
---	---

Simulations of Computer, Telecommunications, Control and Social Systems

Babakina N.A., Kolesnikov M.P., Shkodyrev V.P. <i>An Environment Representation Based on Vector Dynamical Structures in Mobile Robotics Systems</i>	19
--	----

Software of Computer, Telecommunications and Control Systems

Nikulina E.A. <i>Point-Spread Function Computation under the Conditions of Birefringence and Partial Coherence of Lighting using GPUs</i>	33
--	----

Hardware of Computer, Telecommunications and Control Systems

Nikitin A.B., Khabitueva E.I. <i>A Surface Mount Chip Resistor Microwave Model</i>	39
---	----

Information, Control and Measurement Systems

Khristodulo O.I., Salimzynov I.F., Gareeva N.R. <i>The Development of an Information System for Positioning Technogenic Hazard Objects Using Fuzzy Logic</i>	47
---	----

Malykhina G.F., Kislitcina I.A. <i>Measurement of Motion Characteristics Using Neural Networks</i>	59
---	----

Intellectual Systems and Technologies

Sokolov K.V., Timofeev D.A., Samochadin A.V. <i>Extraction of Business Process Models from Texts</i>	69
---	----

Korelin V.N., Blekanov I.B., Sergeev S.L. <i>Clustering of the External WEB Environment of Universities Using a Modified LSH Algorithm</i>	79
---	----

Khassina E.M., Lomov A.A. <i>Audio Files Compression with the STLS-ESM Method</i>	88
--	----

Information and Telecommunications Technologies in Education

Maslov M.Yu., Nosnitsyn S.M., Samochadin A.V., Loginov K.E. <i>An Architecture of an Enterprise Mobility Management Suite for Education Establishments</i>	97
---	----

DOI: 10.5862/JCSTCS.229.1

УДК 004.925

А.Л. Автюшенко, В.М. Иванов

КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ КОНВЕРТАЦИИ ДВУХМЕРНЫХ ОБЪЕКТОВ В ТРЕХМЕРНУЮ СЦЕНУ

A.L. Avtiushenko, V.M. Ivanov

COMPUTER TECHNOLOGIES FOR CONVERTING TWO-DIMENSIONAL OBJECTS TO THE THREE-DIMENSIONAL SCENE

Отмечена неотъемлемая значимость трехмерной визуальной составляющей в восприятии графической информации, произведен анализ различных существующих технологий конвертации реальных и двухмерных объектов в трехмерную или псевдотрехмерную форму. На базе этого анализа кратко изложена одна из синтезированных технологий преобразования двухмерного изображения в псевдотрехмерную форму на примере проекта «Ожившие картины» (проект выполнялся в сотрудничестве с Государственным Русским музеем, результат в виде анимированной панорамы картины И.И. Шишкина демонстрируется в корпусе Кордегардия Михайловского замка).

3D-ИЗОБРАЖЕНИЕ; ТРЕХМЕРНАЯ ГРАФИКА; ДВУХМЕРНАЯ ГРАФИКА; ВОСПРИЯТИЕ ИНФОРМАЦИИ; ЗРЕНИЕ; ИЗОБРАЖЕНИЕ; ТЕХНОЛОГИЯ; АНИМАЦИЯ; КОНВЕРТАЦИЯ; 2D-ИЗОБРАЖЕНИЕ.

In this paper, we have identified the crucial importance of the three-dimensional component in the visual perception of graphical information. We have also analyzed various existing technologies of converting two-dimensional and real objects to a three-dimensional or a pseudo three-dimensional form. One of the adapted conversion technologies is summarized on the basis of this analysis. This new technology is described on the example of the «Revived Pictures» project (a real project executed in collaboration with the State Russian Museum, the result of which is a panorama of an animated picture of Ivan Shishkin demonstrated in the Corps de Garde pavilion of the Mikhailovsky Castle).

3D-IMAGE; 3D-GRAPHICS; 2D-GRAPHICS; INFORMATION PERCEPTION; VISION; IMAGE; TECHNOLOGY; ANIMATION; CONVERSION; 2D-IMAGES.

Последние 15 лет наблюдается невероятный прогресс в создании трехмерных и псевдотрехмерных изображений. Такое активное развитие отрасли компьютерной графики обусловлено активным спросом в различных сферах жизнедеятельности человека. Исследование достижений в разработках трехмерной графики представляет собой одно из наиболее перспективных направлений в психологических, математиче-

ских, сейсмологических, географических и многих других исследованиях. Практически каждый раздел современной науки приобретает уникальную возможность рассматривать и осуществлять свою деятельность, используя преимущества трехмерного восприятия.

В компьютерной графике в целом, и в трехмерной графике в частности, основной информацией являются графические обра-

зы, последовательность этих образов и звук [1]. Основные органы восприятия медиainформации у человека – зрение и слух. Они обладают спецификой, которую необходимо понимать для реализации наиболее приближенного к реальности трехмерного изображения. По информационной теории Д. Марра суть процесса зрительного восприятия лежит в сборе, представлении, обработке и распознавании информации, отражающей свойства наблюдаемого человеком реального мира [2]. Данный алгоритм состоит из следующих шагов: на первом строится эскиз с учетом яркости и геометрических особенностей объектов, далее – 2,5-мерный эскиз глубин и ориентаций поверхностей в системе координат наблюдателя, на последнем выводится трехмерная модель, в координатах объекта которой происходит описание трехмерной структуры и организации формы.

Человек обладает бинокулярным зрением, что дает ему возможность видеть окружающий мир рельефным, ощущать взаимное расположение предметов по глубине [3]. Следовательно, тяга к трехмерному восприятию вполне очевидна.

Ниже будут рассмотрены примеры совмещения и адаптации различных технологий конвертации на примере музейной среды, т. к. именно в этой области необходим постоянный приток современных медиатехнологий, а также она позволяет в ограниченные сроки отследить эффективность и получить обратную связь (благодаря большому количеству вовлеченных участников).

Анализ технологий конвертации объектов в трехмерную форму

Для формирования наиболее полной картины необходимо рассмотреть уже существующие технологии конвертации, их преимущества и недостатки. В современной компьютерной графике трехмерное изображение рассматривается как комбинация объемной полигональной сетки и натянутой на нее текстуры, описывающей цвет изображения [4]. Отсюда становится ясно, что если целью проекта становится создание трехмерной сцены, имитирующей реаль-

ность, то необходимо, во-первых, каким-то образом получить или сформировать данные о полигональной сетке (структуре реальных объектов), а, во-вторых, получить или сформировать текстуры, повторяющие реальные свойства объектов.

Трехмерное сканирование реальных объектов на первый взгляд дает возможность сгенерировать описанные выше элементы. Основной плюс подобного подхода в автоматизации процесса получения полигональной сетки и текстур. Однако стоимость технического оснащения крайне высока, что сдерживает развитие технологии [4]. Также сдерживающим фактором можно назвать несовершенство итоговой 3D-модели (результат может содержать большое количество «артефактов» и искажений).

Всего можно выделить два направления в сканировании реальных объектов: *контактное* и *бесконтактное*.

Бесконтактные методы разделяются на три основных направления:

- 1) технология, основанная на использовании стереоизображений;
- 2) технология, использующая структурированный свет;
- 3) лазерная технология [5].

Алгоритм первой технологии состоит в фотографировании объекта с различных ракурсов и воссоздании трехмерной модели на основе фото. Подобный подход терпит в точности из-за того, что достаточно сложно организовать точное соответствие для пикселей на изображениях, полученных с двух камер, в бесструктурных областях объекта.

Сканирование при помощи структурированного света позволяет с помощью проецирования на реальный объект сетки из линий или точек отследить их искажения на конечных изображениях. Однако процесс является трудоемким и обладает рядом технологических ограничений. К примеру, для воссоздания модели объекта, части которого имеют различную детализацию, необходимо вручную добавлять дополнительные линии или точки в областях, где количество деталей больше, чем на основной модели.

При лазерной технологии применяют



как лазерные датчики, так и цифровую фототехнику [5]. Текстуры для объектов генерируются автоматически. Как и в предыдущем случае технология трудоемка и требует больших временных затрат для оценки и корректировки полученного результата, однако ее несомненным достоинством является возможность конвертировать в трехмерный формат крупные объекты, такие как здания.

В основе *контактных методов* сканирования — высокочувствительный щуп (сенсор), изучающий контуры объекта и выстраивающий его трехмерную копию на экране. Технология проста в использовании, обладает высокой точностью, результат не зависит от освещения. Однако затруднительно применять ее на объектах со сложной геометрией, больших объектах и на объектах, доступ щупа к которым затруднен или невозможен. Также способ крайне затратен по времени и нет возможности захватить данные о текстуре объекта. Область применения — в промышленности.

Следующей технологией, позволяющей осуществлять конвертацию реальных объектов в трехмерные, является *фотограмметрия*. Технология дает возможность интерпретации пространственной информации в заданной системе координат и измерения координат отдельных точек этой модели [6]. Суть технологии фотограмметрии заключается в фиксировании реального объекта на нескольких фотографиях, выполненных с разных ракурсов, обработке полученных снимков с помощью специализированного программного продукта, самостоятельно формирующего связи между изображениями и создающего основу для трехмерного объекта. Преимущество метода заключается в возможности передать мельчайшие неровности и детали объекта (трещинки, пятна, размытия), не используя при этом дорогостоящее оборудование для 3D-сканирования. Методы фотограмметрии также позволяют реконструировать объекты крупных форм, исследовать сооружения в динамическом состоянии [7]. Однако данный способ практически непригоден для конвертации объектов с отражающей поверхностью или объектов из стекла.

Выше рассмотрены всевозможные автоматические и полуавтоматические методы конвертации объектов. Все они обладают теми или иными недостатками и требуют «ручной» доработки. Некоторые объекты в принципе невозможно воспроизвести с помощью подобных технологий (особенно если объект не существует в реальности, либо существует как прототип), в таком случае инженеры прибегают к помощи технологий «ручного» моделирования. Выделяют два основных направления последних: *полигональное моделирование* и *параметрическое*.

Полигональное моделирование — технология трехмерного моделирования, основанная на использовании «полигонов» — многоугольников с тремя или четырьмя вершинами. Каждый полигон или все вместе имеют свой цвет или текстуру и располагаются в виртуальном трехмерном пространстве. Соединение группы таких полигонов позволяет смоделировать практически любой объект (чем больше детализация предполагаемого объекта, тем больше необходимо смоделировать полигонов для отображения деталей). Основным недостатком такого метода можно назвать необходимость объектов быть составленными из крошечных плоских поверхностей, которые должны иметь сколь угодно малый размер, иначе края объектов будут иметь огранный вид, невозможно будет отобразить плавный переход. Из этого следует, что плотность полигонов для фотореалистичного отображения объекта, как правило, должна быть крайне большой, что усложняет и сам процесс моделирования, и итоговый вес модели.

Основная идея трехмерного *твердотельного (параметрического) моделирования* — в проектировании с использованием параметров элементов модели и соотношений между этими параметрами. Параметризация позволяет в режиме реального времени изменять характеристики объекта. Чаще всего применяется в промышленности, т. к. дает возможность получать на выводе прототипы реальных объектов с реальными параметрами, которые можно использовать как исходные значения для различных

станков (например, фрезерных). Метод используется в промышленности, достаточно сложен для освоения (необходимо обладать высоким уровнем знаний или/и быть профессиональным инженером в необходимой области).

Последняя технология – *съемка стерео-видео* крайне сложна технически и базируется на максимально точно откалиброванных и настроенных камерах. Активному развитию технологии также мешает отсутствие четких представлений о том, как формируется ощущение глубины пространства в зрительной системе человека [8]. Из-за данного факта разработчики вынуждены каждый раз опытным путем определять, возможно ли получить желаемый стереоскопический эффект при съемке данной сцены, а если да, то при каких параметрах [9]. Аналогичные проблемы возникают при конвертации двухмерного видеоряда в трехмерный, однако в таком случае процесс усложняется еще и тем, что перенять плохо конвертирующиеся сцены возможности нет.

Принципиальной особенностью наблюдения трехмерных сцен и изображений, которая отличает их от наблюдения двухмерных изображений, является возникновение ощущения глубины за счет бинокулярного зрения [9]. Здесь важно понимать и четко представлять все механизмы, принимающие участие в формировании глубины пространства и оценки расстояния до наблюдаемых объектов, т. к. для каждой конкретной ситуации инженеру приходится опытным путем определять условия получения трехмерного материала и алгоритмы конвертации.

Практическое применение сочетания различных компьютерных технологий конвертации двухмерных объектов в трехмерную сцену

Исходя из всего перечисленного выше, всех достоинств и недостатков технологий конвертации разработана технология *реконструкции трехмерных сцен с помощью проецирования двухмерных изображений на трехмерные или двухмерные объекты*. Основная идея технологии заключается в том, чтобы использовать исходное плоское изобрае-

ние и воссоздать на его основе трехмерную сцену, руководствуясь методиками проецирования изображений на плоские или объемные поверхности и располагая их в пространстве трехмерной сцены.

Разработанная технология несет в себе аспекты методик полигонального моделирования, фотограмметрии и базируется на методике разбиения плоских изображений и послойной реконструкции двухмерного пространства. Основная особенность проекта состоит в том, что послойная реконструкция, как правило, используется для реализации примитивных одноплановых плоскостных анимаций, однако в данной работе с ее помощью реконструируется реалистичное изображение с несколькими планами.

Описанный выше анализ технологий приводит к несомненному выводу о том, что на текущий момент инженер компьютерной графики располагает огромным инструментарием для конвертации двухмерных изображений в трехмерные, для воссоздания реальности на компьютере и т. п., однако, как правило, каждая из этих технологий носит узкоспециализированный характер и не всегда применима для решения конкретных задач, но, в общем случае, требует значительных финансовых затрат.

В рамках настоящей статьи рассмотрены основные механизмы конвертации и различные технологии создания трехмерного изображения на основе реальных объектов. В качестве примера адаптированной технологии приведена конвертация живописного полотна в трехмерную анимированную сцену.

Объектом конвертации выступает панорама на 360° (рис. 1), основанная на шести оригинальных произведениях художника И.И. Шишкина (рис. 2). Анимированный результат конвертации данной плоской панорамы размещен в корпусе Кордегардия на территории Михайловского замка в Санкт-Петербурге. Руководство музея ставило задачу из плоского статичного объекта с минимальными видоизменениями создать иллюзию анимированного пространства, где каждый из объектов сцены подчиняется физическим законам природы, и про-



Рис. 1. Итоговая панорама на основе шести оригинальных работ И.И. Шишкина



Рис. 2. Шесть оригинальных произведений И.И. Шишкина

исходит некоторое действие (перемещение крон деревьев, вызываемое ветром, полет птиц, движение воды и т. п.). Для решения данных задач исследован исходный объект и выделены технологии конвертации, которые могут обеспечить подходящий результат.

Так как объект (картины художника) является плоским, информацию о пространстве невозможно получить ни одним из способов сканирования или фотограмметрии. Съемка стереовидео реальных объектов также не может привести к необходимому результату, т. к. пейзажи на полотнах достоверно воспроизвести невозможно. В данном случае приемлемо лишь использование технологий полигонального моделирования в сочетании с проецированием на воссозданные полигональные объекты.

Формат исходного изображения 10190×1150 пикселей. Формат обусловлен особенностями формирования бесшовных круговых панорамных изображений — лентообразных картин с круговым обзором. Панорамы создают иллюзию реального пространства, окружающего зрителя в полном круге горизонта, и всегда имеют соотноше-

ние сторон один к трем и более (для обеспечения угла обзора, близкого к человеческому взору). Разрешение 72 ppi. Выбор разрешения обусловлен особенностями техники вывода изображений: графика, которая изначально планируется для вывода на экран монитора или проектора, как правило, не требует большего разрешения, а любая избыточная информация грозит увеличением времени обработки информации.

В качестве объектов проецирования выбраны фрагменты живописных полотен, такие как деревья и ветки деревьев, детали переднего плана, камни и четко обозначенные растения. В качестве субъектов проецирования созданы полигональные плоскости, расположенные в пространстве исходя из их предполагаемого размещения в реальном прототипе картины (рис. 3).

На первом этапе создается трехмерная сцена и основная полигональная плоскость (для того чтобы впоследствии натянуть на нее изображение панорамы). Размещение плоскости в сцене не принципиально, т. к. значимым параметром является лишь положение камеры относительно объекта (плоскости картины). Для удобства дальнейших

расчетов основной объект можно располагать в центре координат вертикально. Параметры полигональной плоскости крайне важны, т. к. в случае если пропорции плоскости не будут точно совпадать с пропорциями исходного изображения, текстура исказится. В нашем случае пропорции плоскости должны равняться соотношению 10190/1150 (соотношение сторон исходной панорамы). Далее фронтально по отношению к плоскости размещена ортографическая камера. Важно, чтобы в настройках вида камеры соотношения сторон также совпадали с 10190/1150, а размер окна вида был 10190×1150 пикселей.

Здесь также необходимо связать полигональный объект (будущую плоскость панорамы) с ее текстурой (изображением панорамы).

На последнем этапе устанавливается освещение в сцене. Это необходимо для того, чтобы точно воспроизвести тени от двигающихся объектов. На данном этапе анализируется положение теней на оригинальном изображении и реконструируется положение точечного источника рассеянного света в пространстве. В любом случае, после данного этапа проводится серия тестов с целью убедиться, что источник света отбрасывает тени, аналогичные теням на исходном плоском изображении. Если это не так, положение источника в пространстве корректируется, и тесты возобновляются.

Говоря о тестах, необходимо отметить, что их проведение сводится к тому, что изображение с камеры рендерится (визуализируется) и сопоставляется в редакторе двухмерной графики с изображением-оригиналом. Очень удобна при этом техника работы со слоями, где верхним слоем является полученное изображение-рендер, а нижним — оригинал. При изменении режима смешивания для верхнего

изображения на режим difference просто отследить различающиеся элементы. Данный режим представляет идентичные пиксели как черные, похожие — как темно-серые, а противоположные — как инвертированные. Подобный подход позволяет отследить малейшие неточности еще в процессе работы и избежать глобальных несоответствий. Тесты рекомендуется проводить после каждого последующего шага.

Следующий этап начинается в тот момент, когда пространство исходного изображения разделено по планам (передний, средний и задний). Происходит поиск фрагментов изображения, потенциально пригодных для последующего анимирования.

Затем для каждого из планов (в данном случае — неба, земли и леса) в редакторах двухмерной графики создаются фоны (бэграунды) на прозрачном фоне (вся лишняя информация, не принадлежащая к тому или иному плану, заменяется на прозрачные пиксели, а области, где будут располагаться заанимированные объекты, заретушируются). Каждый фон сохраняется как текстура с разрешением, аналогичным разрешению исходной панорамы, в один из форматов файлов растровой графики, который поддерживает прозрачность. Далее в программе создаются дополнительные полигональные плоскости, аналогичные по параметрам исходной, но размещаются они со смещением по оси глубины, соответственно их плану (задний план — текстура неба, средний — лес, передний — земля). Значение смещения здесь также не принципиально, т. к. камера, через которую будет выводиться окончательный результат, ортографическая.

После окончания подготовительной работы стартует этап ротоскопирования. На этом этапе все крупные фрагменты «вырезаются» с помощью редакторов двухмерной



Рис. 3. Пример текстуры одного из планов (задний план неба и земля)



Рис. 4. Вырезанные фрагменты панорамы, которые впоследствии будут анимироваться

графики или программ, предназначенных для композитинга (например, The Foundry Nuke), из исходного изображения панорамы (крупными в данном случае считаются фрагменты, находящиеся на одном из передних планов, имеющие четкие очертания и являющиеся самостоятельными единицами). Каждый глобальный объект холста, который в реальности был бы отдельной единицей, становится самостоятельным изображением (рис. 4).

Крайне удобно осуществлять упомянутый выше процесс в программах композитинга, а не в программах двухмерной графики, т. к. результирующий объект можно откорректировать в любой момент.

На следующем этапе каждое из созданных изображений сохраняется аналогично текстурам планов в один из форматов растровой графики, поддерживающий прозрачность, и натягивается в качестве текстуры на полигональную плоскость внутри сцены (соотношение сторон плоскости совпадает с соотношениями сторон каждой из текстур). Плоскости также размещаются внутри пространства сцены со смещением по оси глубины, соответственно их плану на исходной картине (положение плоскостей по другим осям совпадает с местонахождением их объектов-прототипов).

Далее, с целью произвести у итогового продукта реалистичную анимацию, полученные на предыдущем этапе изображения разделяются на подобъекты. Например, у каждого дерева выделяются характерные ветви, пригодные для анимации (при невыполнении данного пункта существует риск получить в результате гротескный резуль-

тат, при котором все дерево в целом будет раскачиваться на ветру, а ветви дерева — нет).

Необходимо уточнить, какие именно подобъекты считаются «пригодными для анимации». Во-первых, подобъект должен быть четко очерчен в плоскости изображения, т. е. не иметь размытых контуров. Во-вторых, размеры полученного изображения должны быть не меньше $0,5 \times 0,5$ мм при разрешении в 72 ppi (т. е. 36 пикселей по ширине и высоте), иначе существует риск проделать большой объем работы и не получить видимого результата (т. к. столь малое число пикселей не может содержать визуально-ощутимой для человеческого глаза информации).

После окончания основной работы по разбиению и формированию объектов сцены проводится коррекция промежуточного результата (ретушь в графическом редакторе): с фона основной картины-подложки убираются элементы, которые будут далее анимироваться. Данный шаг выполняется во избежание появления анимированного предмета в статичной форме на заднем плане, под самим анимируемым объектом при его смещении (рис. 5).

На следующем этапе анализируются реальные прототипы объектов и характер их движений в природе с целью написания математической функции (как правило, несколько), обеспечивающей похожие на реальные колебания. Наиболее распространенным примером управления скоростью является движение с плавным входом/выходом. При таком способе управления скоростью происходит плавное достижение

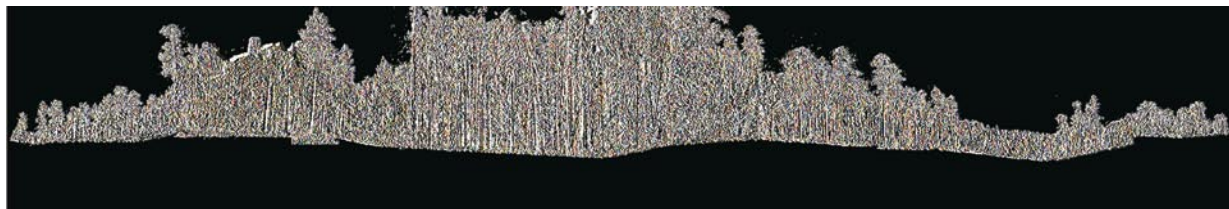


Рис. 5. С изображения-подложки убираются все объекты с анимацией

максимальной скорости и плавное торможение [10]. Для этого функцию скорости часто выбирают синусоидой. В данном случае выбрана синусоидальная функция, отвечающая за скорость качаний и их частоту. Случайный элемент (второй множитель) необходим для того, чтобы все объекты не двигались абсолютно одинаково:

$$V = \sin\left(\frac{frame}{2_0} + 2\right) * random\left(\frac{frame}{1_s}\right) * 3.$$

Преимущество данной формулы также заключается в том, что аниматор может в любой момент с легкостью ускорить или замедлить движение, изменив в формуле последний множитель (чем он больше, тем движение активнее, чем меньше – тем медленнее).

После применения анимации по формуле к каждому из потенциальных субъектов анимации, тестируется схожесть воссозданного пространства и плоского оригинала. Для этого выводится секвенция из кадров (соответственно длительности анимационного видеоролика), ключевые кадры которой последовательно сопоставлены с их оригинальным изображением в редакторе двухмерной графики (аналогичный процесс был описан ранее для тестирования светотеневого решения).

Значительные несовпадения устраняются. Затем начинается заключительный этап: создается звуковое сопровождение и добавляются детали, полученные в результате дополнительной видеосъемки (полет птиц, пыль, дождь и т. п.).

Таким образом, итоговая сцена состоит из набора плоскостей, расположенных в пространстве с небольшим сдвигом по оси глубины (соответственно плановости оригинала). На каждую плоскость натяну-

ты текстуры объектов или групп объектов с использованием масок прозрачности. На все это множество плоскостей направлена ортографическая камера с пропорциями сторон, равными пропорциям изображения-оригинала. В сцене также присутствует рассеянный источник света для имитации отбрасываемых от движущихся объектов теней.

Технология, описанная выше, находится на стадии разработки и не может называться универсальным средством для решения проблем в сфере конвертации объектов в трехмерный формат, однако она успешно применялась на практике в различных проектах (например, с помощью аналогичной технологии выполнен другой музейный проект по картине Г.Г. Чернецова «Парад на Царицыном лугу» (заказчик – Государственный Русский музей) или проект в сотрудничестве с Государственным Дарвиновским музеем города Москвы – «Ожившие места Лосиногостовского острова») и позволяла удешевить и ускорить производственный процесс, являясь предпосылкой к формированию нового оптимального способа для конвертации реального или графического материала в трехмерный формат.

Большое разнообразие современных методов получения информации о трехмерной составляющей лишь в малой мере облегчает процесс конвертации двухмерной графики в трехмерную. Каждый проект имеет в своей основе индивидуальные задачи, цели, финансовые и иные ограничения, в соответствии с которыми необходимо корректно выбирать технологию или комплекс технологий конвертации. Задачи, которые решаются с помощью средств компьютерной графики, нередко требуют от разработчика совершенно новых способов их реше-



ния, что влечет за собой создание новых технологий.

На основе проведенных исследований можно сделать вывод о том, что для конвертации реальных объектов умеренной сложности (например, объектов малых форм: архитектурных артефактов, предметов быта, скульптур и пр. с размерами до двух метров по всем измерениям) наиболее эффективно использовать технологии фотограмметрии в сочетании с технологиями бесконтактного сканирования. Доработать результат лучше всего с помощью технологий полигонального моделирования. Если объект достаточно большой по размеру (здания, крупные элементы архитектуры), приемлемо прибегать к помощи лазерного сканирования и технологий фотограмметрии. В случае если объект, который следует конвертировать, необходим для целей промышленности и производства (различного рода детали и фрагменты установок), лучшим способом получения трехмерной информации будет метод контактного сканирования в сочетании с параметрическим моделированием. Вариант, когда трехмерную составляющую необходимо получить из изначально плоских двухмерных объектов, предполагает сочетание нескольких методик конвертации: проецирование двухмерной графики на трехмерные или плоские объекты, полигональное моделирование и (если возможно) съемка стереовидео.

Технология, представленная на примере проекта-панорамы, сочетает в себе многие плюсы из традиционных технологий конвертации, не имея при этом их значительных минусов: она выгодна в финансовом плане (ее полностью в состоянии выполнить один квалифицированный инженер, используя доступную и недорогостоящую технику), сравнительно не затратна по времени (в зависимости от целей и задач конкретного проекта работа по конвертации может занимать от одного дня до месяца), позволяет работать с разнообразным исхо-

дным материалом (от полотен художников до фотографий и словесных концепций) и позволяет воплощать практически любой уровень детализации (зависит лишь от качества оригинального изображения или фотографии).

Используемые технологии послойной реконструкции двухмерного изображения в условиях трехмерной сцены и проецирования фрагментов оригинальной графической информации на полигональные плоскости являются новыми. Широкий опыт их применения еще не накоплен. Все проекты, выполненные на базе данной методики, – нетиповые, первые в своем роде, поэтому не существует какого-либо прототипа для их сравнения и анализа. Основным критерием оценки выполненных по данной технологии работ послужило экспертное мнение специалистов Государственного Русского музея в Санкт-Петербурге (для приведенного выше проекта и реконструкции картины Г.Г. Чернецова «Парад на Царицыном лугу») и Государственного Дарвиновского музея в Москве (для проекта «Ожившие картины Лосино острова»), где и выставлены медиапроекты. Таким образом, все полученные в ходе исследований результаты имеют практическую ценность для улучшения аналогичных проектов в будущем.

В заключение можно заметить, что создание новых технологий всегда происходит на базе уже существующих. Поэтому залогом достижения эффективного результата, который будет восприниматься наиболее корректным образом, становится успешный выбор и адаптация базовых технологий конвертации. Технология, описанная в данной статье, имеет хорошие перспективы, т. к. может успешно сочетаться с любым из обозначенных в работе методов конвертации двухмерной составляющей в трехмерную. Методику можно использовать и как базовую, и как аддитивную для создания проектов любого уровня сложности и качества.

СПИСОК ЛИТЕРАТУРЫ

1. Седов Е.А. Информационно-энтропийные свойства социальных систем // Заседание Междисциплинарного семинара по изучению цивили-

лизационных кризисов. 1993.

2. Марр Д.К. Зрение. Информационный подход к изучению представления и обработ-

ки зрительных образов. Москва: Радио и связь, 1987. 401 с.

3. **Майзель С.О.** Свет и зрение. Изд-во ГТТИ, 1932. 124 с.

4. CSD Центр быстрого прототипирования // Технология и перспективы 3D сканеров [электронный ресурс]/ URL: <http://print3d.ru/index.php/zakaz-prototipa/zakaz-skanirovaniya.html> (дата обращения: 28.06.2015).

5. **Красильников Н.Н.** Метод получения 3D-изображений, основанный на диффузном отражении света сканируемыми объектами // Информационно-управляющие системы. 2009. № 6. С. 7–11.

6. **Павленко А.В.** Разработка методики создания фотограмметрических моделей местности по аэрокосмическим снимкам: дис. ... канд. техн. наук. Новосибирск, 2006.

7. **Лобов М.И., Переварюха А.Н., Чирва А.С.**

Применение наземной фотограмметрии и лазерного сканирования для исследования динамического состояния мачтовых сооружений // Вестник Донбасской национальной академии строительства и архитектуры. Технология, организация, механизация и геодезическое обеспечение строительства. 2010. № 3(83).

8. **Акимов Д.А.** Выделение полупрозрачных частиц переднего плана в видео на основе анализа резкости кадра // Сб. тезисов XX междунар. науч. конф. Ломоносов-2013. Москва: МГУ им. М.В. Ломоносова, 2013. 156 с.

9. **Красильников Н.Н.** Механизмы и погрешности восприятия 3D-пространства зрительной системой // Информационно-управляющие системы. 2014. № 6. С. 13–18.

10. **Синявский В.** История развития анимации. Обзор основных методов компьютерной анимации. Москва, 2011. 25 с.

REFERENCES

1. **Sedov Ye.A.** *Informatsionno-entropiynnye svoystva sotsialnykh sistem [Information and entropy properties of social systems]*, 1993. (rus)

2. **Marr D.K.** *Zreniye. Informatsionny podkhod k izucheniyu predstavleniya i obrabotki zritelnykh obrazov [Vision Information approach to the study of representation and processing of visual images]*. Moscow: Radio i svyaz Publ., 1987, 401 p. (rus)

3. **Mayzel S.O.** *Svet i zreniye [Light and vision]*. GTTI Publ., 1932, 124 p. (rus)

4. CSD *Tsentr bystrogo prototipirovaniya. Tekhnologiya i perspektivy 3d skanerov [Technology and prospects of 3D scanners]*. Available: <http://print3d.ru/index.php/zakaz-prototipa/zakaz-skanirovaniya.html> (Accessed: 28.06.2015). (rus)

5. **Krasilnikov N.N.** *Metod polucheniya 3D-izobrazheniy, osnovanny na diffuznom otrazhenii sveta skaniruyemyimi obyektami [A Method of 3D Scanning Based on Scanned Objects' Light Radiosity]*, *Informatsionno-upravlyayushchiye sistemy [Information and Control Systems]*, 2009, No. 6, Pp. 7–11. (rus)

6. **Pavlenko A.V.** *Razrabotka metodiki sozdaniya fotogrammetricheskikh modeley mestnosti po aerokosmicheskim snimkam [Development of the method of creation of photogrammetric terrain models on aerospace images]*. Novosibirsk, 2006. (rus)

7. **Lobov M.I., Perevaryukha A.N., Chirva A.S.**

Primeneniye nazemnoy fotogrammetrii i lazernogo skanirovaniya dlya issledovaniya dinamicheskogo sostoyaniya machtovykh sooruzheniy [The use of photogrammetry and terrestrial laser scanning for the study of the dynamic state of mast structures], *Vestnik Donbasskoy natsionalnoy akademii stroitelstva i arkhitektury. Tekhnologiya, organizatsiya, mekhanizatsiya i geodezicheskoye obespecheniye stroitelstva [Bulletin of Donbas National Academy of Civil Engineering and Architecture. Technology, organization, mechanization and geodetic support of construction]*, 2010, No. 3(83). (rus)

8. **Akimov D.A.** *Vydeleniye poluprozrachnykh chastits perednego plana v video na osnove analiza rezkosti kadra [Isolation translucent particles foreground in the video based on the analysis of field frame]*. Moscow: MGU imeni M.V. Lomonosova Publ., 2013. 156 p. (rus)

9. **Krasilnikov N.N.** *Mekhanizmy i pogreshnosti vospriyatiya 3D-prostranstva zritelnoy sistemoy [Mechanisms and Inaccuracies of 3D Spatial Perception by Visual System]*, *Informatsionno-upravlyayushchiye sistemy [Information and Control Systems]*, 2014, No. 6, Pp. 13–18. (rus)

10. **Sinyavskii V.** *Istoriya razvitiya animatsii. Obzor osnovnykh metodov kompyuternoy animatsii [The history of animation. Overview of the main methods of computer animation]*. Moscow, 2011, 25 p. (rus)

АВТЮШЕНКО Алла Леонидовна – инженер I категории, ассистент кафедры инженерной графики и дизайна Института металлургии, машиностроения и транспорта Санкт-Петербургского политехнического университета Петра Великого.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: allakadabralla@mail.ru



AVTIUSHENKO Alla L. *Peter the Great St. Petersburg Polytechnic University.*
195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.
E-mail: allakadabralla@mail.ru

ИВАНОВ Владимир Михайлович — *заведующий кафедрой инженерной графики и дизайна Института металлургии, машиностроения и транспорта Санкт-Петербургского политехнического университета Петра Великого, доктор физико-математических наук.*

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: ivm@imop.spbstu.ru

IVANOV Vladimir M. *Peter the Great St. Petersburg Polytechnic University.*
195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.
E-mail: ivm@imop.spbstu.ru



FEDERATED CONFERENCE ON COMPUTER SCIENCE AND INFORMATION SYSTEMS

Gdańsk, Poland, 11-14 September, 2016

Call for Papers:

The FedCSIS Events provide a platform for bringing together researchers and practitioners to present and discuss ideas, challenges, and new solutions in computer science and information systems. Topics of interest are defined by Events constituting FedCSIS and listed on www.fedcsis.org

The papers should be submitted to a chosen Event by April, 18, 2016 using the FedCSIS submission system www.fedcsis.org/hotcrp/

Accepted and presented papers will be submitted to the IEEE Xplore Digital Library to be published in proceedings entitled "2016 Federated Conference on Computer Science and Information Systems (FedCSIS)". Because the IEEE proceedings will be published under nonexclusive copyright, the Events' organizers will endeavor to arrange quality journals, edited volumes, etc. and will invite extended and revised papers for post-conference publications.

Important Dates:

Submission of Event Proposals: **October 30, 2015**

Paper submission: **April 18, 2016**

Author Notification: **June 13, 2016**


Final submission and registration: **July 4, 2016**

Conference date: **September 11-14, 2016**



e-mail: secretariat@fedcsis.org
www.fedcsis.org





Моделирование вычислительных, телекоммуникационных, управляющих и социально-экономических систем

DOI: 10.5862/JCSTCS.229.2

УДК 004.896:004.925.8

Н.А. Бабакина, М.П. Колесников, В.П. Шкодырев

ПРЕДСТАВЛЕНИЕ ОКРУЖАЮЩЕГО ПРОСТРАНСТВА НА ОСНОВЕ ВЕКТОРНЫХ ДИНАМИЧЕСКИХ СТРУКТУР В МОБИЛЬНЫХ РОБОТОТЕХНИЧЕСКИХ СИСТЕМАХ

N.A. Babakina, M.P. Kolesnikov, V.P. Shkodyrev

AN ENVIRONMENT REPRESENTATION BASED ON VECTOR DYNAMICAL STRUCTURES IN MOBILE ROBOTICS SYSTEMS

Рассмотрен подход к созданию геометрических векторных моделей для описания окружающего пространства. В совокупности данные модели образуют карту местности, на основе которой могут решаться задачи локализации робота, а также задачи управления движением. Также векторные модели предоставляют дополнительную возможность распознавания объектов по их геометрическим характеристикам. Это достигается благодаря инвариантным характеристикам моделей относительно движений объекта (робота) в пространстве. Функциональная особенность метода связана с обработкой вторичных данных на основе кусочно-линейной аппроксимации контуров, выделяемых из образов окружающего пространства. Такой подход допускает применение сенсорных систем, основанных на различных физических принципах работы. Описаны алгоритмы последовательного уточнения моделей объектов и адаптивной статистической оценки результатов, которые гарантируют достоверность выстраиваемой карты.

ТЕХНИЧЕСКОЕ ЗРЕНИЕ; ЛОКАЛИЗАЦИЯ; SLAM; ПРЕДСТАВЛЕНИЕ ОКРУЖАЮЩЕЙ СРЕДЫ; СОСТАВЛЕНИЕ КАРТЫ; ГЕОМЕТРИЧЕСКАЯ МОДЕЛЬ ОБЪЕКТА.

The article discusses an approach to creating vector geometric models for environment representation. In conjunction these models form a hierarchical map of some area. The problems of a mobile robot's localization and motion control can be solved using these structures. Moreover, vector models provide an additional opportunity for recognizing objects by their geometric characteristics. This is achieved by the invariant features of the model with respect to the motions of the object (robot) in space. A functional feature of the method is processing the secondary data based on a piecewise linear approximation of the edges allocated from the images of the surrounding space. This approach allows to use sensors based on different physical principles of operation. The algorithms of successive refinement of object models and adaptive statistical evaluation of the results are described. These algorithms guarantee the precision of the designed map.

MACHINE VISION; LOCALIZATION; SLAM; ENVIRONMENT REPRESENTATION; ENVIRONMENT MAPPING; GEOMETRIC MODEL OF THE OBJECT.

Разработка автономных мобильных систем — одна из наиболее актуальных задач

современной робототехники. Наглядным примером может служить ликвидация по-

следствий аварии на атомной электростанции Фукусима-1 (г. Окума, Япония). Оснащенность роботами, способными автономно работать в экстремальных условиях радиационной катастрофы, оказалась фактически нулевой [1]. Более удачными примерами в этом отношении стали системы PackBot 510 и Warrior 710 производства iRobot (США), а также спасательные роботы, разрабатываемые в Германии и Франции [2, 3]. Примечательно, что эти устройства также не являются полностью автономными.

Важнейшая задача при разработке автономной системы управления – проблема полного и достоверного описания окружающего пространства в условиях недостатка априорной информации. Такое описание включает в себя как геометрическое представление мира (составление карты) и локализацию относительно выделенных объектов, так и описание этих объектов с целью их идентификации, классификации, целенаправленного перемещения робота и решения других задач.

Наиболее известным подходом к описанию окружающего пространства в автоматическом режиме являются методы, объединенные в семейство SLAM (Simultaneous Localization and Mapping – Одновременная локализация и составление карты). Но даже наиболее совершенные методы этой группы (например, DP-SLAM, FastSLAM 2.0, EKF-based SLAM и др.) имеют ряд недостатков. Самым серьезным из них является требование к статичности окружающего пространства [4, с. 871–886], тогда как современная автономная роботизированная система должна функционировать в частично наблюдаемом, динамически изменяющемся, стохастическом окружении [5]. Дополнительные ограничения методов SLAM связаны с критическим возрастанием объема данных при увеличении исследуемой зоны или точности. В то же время, выделенные объекты фактически не имеют описания [4, 6]. Более совершенными являются методы visual SLAM, основанные на использовании систем технического зрения, но сложность вычислительных процедур и высокое энергопотребление ограничивают применение

этих технологий в мобильной робототехнике [7, с. 546–550].

Таким образом, проблема полного и инвариантного описания окружающего пространства не имеет наилучшего решения и является одной из наиболее актуальных в мобильной робототехнике.

1. Геометрическое представление окружающего пространства

Представление физических объектов. В работах [8, 9] предложен метод геометрического описания исследуемого пространства на основе кусочно-линейной аппроксимации выделенных контуров. В том случае, если первичные данные являются изображениями с одной или нескольких камер, они могут быть представлены в виде набора гранично-ориентированных проекций на сенсорные элементы робота. Аналогичный подход можно использовать и для данных, полученных в результате радиолокационного, ультразвукового или лазерного сканирования [5, 8, 9]. При использовании средств видеофиксации наиболее достоверный результат предоставляют методы, основанные на регистрации фазово-частотных изменений отражаемого света (например, [10]). Эффективность таких алгоритмов в основном связана с незначительным влиянием изменения яркости изображения в области границ на результат их определения.

Математически полученные данные могут быть описаны как полиэдры, представляющие собой объединение простейших симплексов пространства заданной размерности [11, с. 21–34]. На основании этого свойства в работах [8, 9] показано, что наблюдаемый физический объект H может быть представлен как набор \mathcal{H} 1-симплексов (отрезков) следующим образом:

$$L : H \rightarrow \mathcal{H} = \sum_k \sigma_k, \quad (1)$$

где σ_k – k -й отрезок с координатами, заданными в относительной системе координат робота.

Пусть множество точек $P \subseteq E$ принадлежат поверхности объекта H ; E – глобальное n -мерное пространство. Тогда отрезок с началом в точке $p_1 \in P$ и с концом в точке $p_2 \in P$

можно задать как множество точек:

$$\begin{aligned} \sigma(p_1, p_2) = \\ = \{\alpha p_2 + (1 - \alpha)p_1 \mid p_1, p_2 \in P, \alpha \in [0, 1]\}, \end{aligned} \quad (2)$$

при этом точка $p \in P$ может быть представлена отрезком с совпадающими концами. Следовательно, образ \mathcal{H} физического объекта H будет иметь следующий вид:

$$\mathcal{H} = \sum_k \sigma_k = \sum_k \sigma(p_1^k, p_2^k). \quad (3)$$

Инвариантность векторной модели. Рассмотрим ситуацию, когда наблюдатель находится в заранее не известной случайной точке пространства относительно исследуемого объекта, имея возможность произвести процедуру наблюдения. Наблюдателем ранее сформирована полная геометрическая модель некоторого объекта. По результатам наблюдения необходимо определить, принадлежит ли наблюдаемая часть тому объекту, для которого сформирована модель.

Для решения поставленной задачи необходимо найти такие величины, которые однозначно определяют заданный объект и являются инвариантными относительно преобразований вращения $g(\varphi)$ масштабирования $g'(\alpha)$. Группа преобразований $SO(\mathbb{R}^3)$ оставляет неизменной евклидову метрику пространства. Следовательно, при действии $g \in SO(\mathbb{R}^3)$ будут сохраняться расстояния и углы геометрического объекта. С другой стороны, действия из группы $O(\mathbb{R}^3)$ сохраняют в общем случае только углы. При заданном множестве действий g' будут сохраняться геометрические пропорции. Рассмотрим теперь свойства модели объекта.

Модель объекта задается как множество отрезков, расположенных друг относительно друга определенным образом (3). Следовательно, одну модель от другой отличает количество отрезков, их параметры и взаимное расположение, и два объекта можно считать равными, если они описаны моделями, состоящими из равных множеств элементов (отрезков). С этой точки зрения можно говорить о «структурной топологии» объекта как критерии инвариантности. Иначе говоря, структурная топология данного множества является основным свойством, позволяющим идентифицировать

объект по его модели и отличать от других. Основными рассматриваемыми характеристиками этой модели в данном случае являются размеры отрезков (сегментов) и их взаимное расположение. Преобразование масштабирования $g'(\alpha)$ сохраняет геометрические пропорции, что приводит к необходимости рассмотрения отношений длин между различными отрезками. Преобразование вращения $g(\varphi)$ сохраняет углы, что дает право производить первичную идентификацию объектов именно по этим характеристикам.

Согласно определению (2), сегмент модели задается посредством двух точек p_1, p_2 , что позволяет однозначно определить угол γ наклона прямой, проведенной через него, относительно заданной оси (например, Ox).

Действие $g(\varphi)$ приведет к изменению углов наклона каждого сегмента, но не изменит их разности:

$$\begin{aligned} \gamma_1 \rightarrow (\gamma_1 - \varphi), \quad \gamma_2 \rightarrow (\gamma_2 - \varphi); \\ \Delta\gamma = (\gamma_2 - \gamma_1) \rightarrow \Delta\gamma' = \\ = [(\gamma_2 - \varphi) - (\gamma_1 - \varphi)] = \Delta\gamma. \end{aligned}$$

Таким образом, угол между двумя сегментами инвариантен относительно поворотов и масштабирования и может использоваться для идентификации объекта по сформированной в результате наблюдения модели. При этом удобно вместо значений величины углов использовать их тангенсы [8].

2. Наблюдение стационарного объекта

Движение наблюдателя относительно неподвижного объекта может быть представлено двумя взаимно независимыми процессами: круговым движением вокруг объекта и изменением расстояния наблюдения. С физической точки зрения движение наблюдателя (робота) может быть перенесено на сам объект в соответствии с принципом относительности. Таким образом, можно говорить о вращении объекта на угол (строго говоря, вектор углов) φ и изменении расстояния между объектом и наблюдателем. Изменение расстояния между объектом характеризует изменение

масштаба образа объекта, получаемого при его наблюдении, с коэффициентом α .

Вращение объекта. Пусть наблюдатель производит наблюдение объекта из позиции P_1 , в результате которого строится геометрическая модель μ_1 . Данная модель, как было сказано выше, описывает лишь видимую часть физического объекта, полностью базируясь на проекциях объекта на сенсорные элементы робота. Поэтому она несет лишь частичную информацию о внешней геометрии наблюдаемого объекта:

$$\mu_1 = \sum_{k=0}^{N_1} \sigma_k^{\mu_1},$$

где $\sigma_k^{\mu_1}$ — отрезки, соответствующие модели μ_1 .

Зададим вращательное движение (действие) наблюдателя вокруг объекта как вращение последнего на некоторый угол $g(\varphi)$. При этом наблюдатель переместится из позиции P_1 в позицию P_2 и будет производить новую процедуру наблюдения, в результате которой получим новую геометрическую модель объекта μ_2 :

$$\mu_2 = \sum_{k=0}^{N_2} \sigma_k^{\mu_2},$$

где $\sigma_k^{\mu_2}$ — отрезки, соответствующие модели μ_2 . При этом в общем случае $N_1 \neq N_2$, а расположение и размеры отрезков $\sigma_k^{\mu_1}$ и $\sigma_k^{\mu_2}$ различны.

На основе набора $\{\mu_1, g(\varphi), \mu_2\}$ синтезируем две новые модели следующим образом. Так как движение g задано и известна исходная модель μ_1 , то при данном действии она перейдет в модель $\mu'_2 = g\mu_1$, где в общем случае действие g на модель μ сводится к действиям на составляющие ее сегменты:

$$g\mu = \sum_k g\sigma_k.$$

Модель μ'_2 характеризует представление в позиции P_2 тех частей наблюдаемого объекта, которые были зафиксированы при нахождении наблюдателя в позиции P_1 . Так как эти части могут быть не наблюдаемы в позиции P_2 , то уточненная модель может быть получена посредством объединения:

$$\tilde{\mu}_2 = \mu_2 \cup \mu'_2 = \mu_2 \cup g\mu_1.$$

Вторая синтезируемая модель μ'_1 полу-

чается с помощью обратного действия g^{-1} :

$$\mu'_1 = g^{-1}\mu_2.$$

Данная модель характеризует представление в позиции P_1 тех частей наблюдаемого объекта, которые были зафиксированы при нахождении наблюдателя уже в новой позиции P_2 . Уточненная модель в данном случае будет следующей:

$$\tilde{\mu}_1 = \mu_1 \cup \mu'_1 = \mu_1 \cup g^{-1}\mu_2.$$

Очевидно, что данные модели должны быть эквивалентны при условии установки соотношения эквивалентности в соответствии с действием g :

$$g\tilde{\mu}_1 = \tilde{\mu}_2.$$

Изменение масштабов объекта. Движение наблюдателя вдоль радиус-вектора на объект приведет к тому, что образ объекта, фиксируемый сенсорными элементами, будет пропорционально изменять свои размеры, т. е. будет происходить масштабирование образа объекта с коэффициентом α . Данное действие может быть задано с помощью линейного оператора в матричной форме как $g'(\alpha) = \alpha E_n$, где $\alpha \in \mathbb{R}^1$ — коэффициент масштабирования, а E_n — единичная матрица размера $n \times n$.

При действии g' структура модели наблюдаемого объекта при переходе из точки наблюдения P_1 в точку наблюдения P_2 не изменяется. Поэтому действия масштабирования в идеальном случае не влияют на уточнение модели объекта. Однако в реальности необходимо учитывать разрешающую способность сенсорных элементов, поскольку это может привести к потере данных при достижении определенного граничного случая. Рассмотрим эту ситуацию подробнее.

Пусть сенсорные элементы робота обладают разрешающей способностью, заданной в виде линейного размера фиксирующего элемента δx . Обозначим в качестве Pg оператор проектирования отрезка σ_k на сенсорный элемент. Тогда потеря данных будет происходить в случае значительного уменьшения длины проекции отрезка по отношению к разрешающей способности сенсорного элемента:

$$\delta x > \alpha |\text{Pr } \sigma_k|.$$

С другой стороны, потеря данных возможна также и при увеличении коэффициента α (приближение). При достаточно большом значении α может произойти исчезновение отрезка σ_k из зоны видимости сенсорного элемента. Как в первом, так и во втором случае все зависит от качества сенсорных элементов наблюдателя.

3. Стохастические модели

Описанный выше подход носит детерминированный характер. Однако в реальной ситуации движение наблюдателя и процедура наблюдения являются стохастическими. Следовательно, необходимо перейти к стохастической модели процедуры наблюдения, а также к стохастической модели движения наблюдателя [5, 8, 9].

При описании отдельного физического объекта была предложена множественная модель с использованием 1-симплексов (3). Предположим, что модель \mathcal{H} является наиболее полной геометрической моделью объекта H , то есть

$$\forall \hat{\mathcal{H}}, \hat{\mathcal{H}} \cup \mathcal{H} \equiv \mathcal{H},$$

где $\hat{\mathcal{H}}$ – идеальная модель объекта (рассматривается в том же базисе, что и \mathcal{H}).

При единичном наблюдении формируется текущая наблюдаемая модель физического объекта μ . При этом даже в случае идеальной процедуры наблюдения об объекте будет получена лишь частичная информация, т. к. μ является подмножеством полной идеальной модели $\mu \subseteq \mathcal{H}$, где эквивалентность возможна только для самых простых объектов.

Тогда последовательность формирования геометрической модели наблюдаемого физического объекта H может быть представлена следующим образом:

$$H \rightarrow \mathcal{H} \rightarrow \mu.$$

Рассмотрим переход $\mathcal{H} \rightarrow \mu$. Для определения процедуры наблюдения выделим оператор, позволяющий получить максимально полную геометрическую модель физического объекта при нахождении в позиции наблюдателя. То есть \mathcal{H} и наблюдатель имеют различные ортогональные реперы, и физи-

ческий объект находится на некотором неизвестном расстоянии $\text{Pr} : \mathcal{H} \rightarrow \hat{\mathcal{H}}$, причем $\hat{\mathcal{H}} \subseteq \mathcal{H}$. Оператор Pr является идемпотентным, т. е. повторное действие не изменит его: $\text{PrPr} = \text{Pr}$, что позволяет использовать его как оператор проектирования.

Теперь рассмотрим оператор наблюдения O , позволяющий получить из идеальной модели $\hat{\mathcal{H}}$ геометрическую модель физического объекта μ :

$$O : \hat{\mathcal{H}} \rightarrow \mu.$$

Случай идеального наблюдения приводит к тому, что $\mu \equiv \hat{\mathcal{H}}$. Но в реальной ситуации необходимо учесть, что

- 1) элементы множества $\hat{\mathcal{H}}$ могут не войти в состав множества μ ;
- 2) процедура наблюдения переведет отрезок $\sigma_i \in \mathcal{H}$ в другой отрезок $\tilde{\sigma}_i \in \mu$.

Рассмотрим первое замечание. При наблюдении ненулевые элементы из $\hat{\mathcal{H}}$ могут стать элементами множества μ , но могут и не попасть в него, что обусловлено различными факторами, влияющими на результат наблюдения. Например, частичное размытие образа не позволит выделить некоторые сегменты. Случай непопадания элемента во множество μ зададим как преобразование в нулевой элемент (вырождение):

$$O(\hat{h}) \rightarrow 0, \hat{h} \in \ker O, \hat{h}, 0 \in \hat{\mathcal{H}},$$

где $\ker O$ – ядро оператора наблюдения O .

Ранее были рассмотрены действия, приводящие к изменению геометрической модели объекта – повороты и масштабирование. Действие поворота g влияет только на оператор проектирования Pr . Действие масштабирования g' влияет только на оператор наблюдения O .

Рассматривая реальную ситуацию, необходимо установить зависимость оператора наблюдения от разрешающей способности сенсорных элементов робота. Эти зависимости оператора O будут индуцированы также и на ядро $\ker O$:

$$O(\alpha, \delta) \rightarrow K(\alpha, \delta) = \ker O(\alpha, \delta).$$

Случайный характер процедуры наблюдения приведет к тому, что множество $K(\alpha, \delta)$ также будет случайным. Пусть задан элемент $h \in \hat{\mathcal{H}}$. Зададим для него вероят-

ность попадания в ядро оператора наблюдения как $P^K(h)$, $h \in (\alpha, \delta)$.

Тогда вероятность попадания элемента h во множество μ :

$$P^\mu(h) = 1 - P^K(h).$$

Функции $P^K(h)$ и $P^\mu(h)$ имеют все свойства функций принадлежности для множеств $K(\alpha, \delta)$ и μ . Совместно с множеством $\hat{\mathcal{H}}$ они задают K и μ как нечеткие множества [12, с. 51].

Рассмотрим теперь второе замечание о стохастичности процедуры наблюдения. Помимо того, что элемент идеальной модели $\hat{\mathcal{H}}$ может попасть в наблюдаемую модель μ с заданной вероятностью, при наблюдении он претерпевает случайные изменения, то есть

$$h \in \hat{\mathcal{H}} \rightarrow \tilde{h} \in \mu, \quad h \neq \tilde{h}.$$

В общем случае это приводит к получению стохастической модели наблюдения μ , такой, что $\mu \cap \hat{\mathcal{H}} = \emptyset$.

Рассмотрим отдельно элементы h и \hat{h} . Так как геометрически они представляют собой отрезки, можно задать невязку в виде

$$\begin{aligned} \Delta h &= \tilde{h} - h = \sigma(\tilde{p}_1, \tilde{p}_2) - \sigma(p_1, p_2) = \\ &= \sigma(\tilde{p}_1 - p_1, \tilde{p}_2 - p_2), \end{aligned}$$

где арифметические операции над точками производятся для каждой из координат по отдельности.

Отрезок \hat{h} является стохастическим, поэтому его свойства переносятся на отрезок Δh ввиду детерминированности h . Форма задания отрезков (1) позволяет считать стохастическими невязки $\Delta p_1 = \tilde{p}_1 - p_1$, $\Delta p_2 = \tilde{p}_2 - p_2$.

Поскольку $p_i \in \mathbb{R}^n$, то и $\Delta p_i \in \mathbb{R}^n$. Отрезок h можно представить как точку в пространстве \mathbb{R}^{2n} . Тогда и отрезок Δh может быть представлен случайной точкой в этом же пространстве.

Пусть каждая координата точки Δp_i , $i = \{1; 2\}$, является случайной величиной с нормальным распределением плотности вероятности:

$$f(x) = \frac{1}{\sqrt{2\pi} |D|^{1/2}} e^{-\frac{x^2}{2D}},$$

где D – дисперсия случайной величины.

Согласно [13, с. 380–385], элемент Δh будет случайным вектором с многомерным гауссовым распределением вероятностей, зависящим от $2n$ переменных:

$$f_{2n}(x) = \frac{|A|^{1/2}}{(2\pi)^{2n/2}} \exp\left\{-\frac{1}{2}(Ax, x)\right\}, \quad (4)$$

где $x = (x_1, \dots, x_{2n})$; $A = R^{-1}$ и R – ковариационная матрица.

Предположим, что, во-первых, все случайные величины, которые задают координаты невязок, являются независимыми. Это приводит к диагональному виду ковариационной матрицы R и невырожденности матрицы A . Во-вторых, все случайные величины имеют нормальное распределение плотности вероятности.

Согласно (4), вероятность попадания сегмента Δh в $2n$ -мерный шар с радиусом r :

$$P_c(r) = \int_{|x| < r} f_{2n}(x) dx.$$

Устанавливая вероятность $P_c(r)$ и зная среднее отклонение, можно решать обратную задачу по определению радиуса $2n$ -мерного шара, в который с указанной вероятностью попадет точка $2n$ -мерного пространства Δh . Таким образом, для определения эквивалентности сегментов наблюдаемой модели μ и сформированной модели \mathcal{H} объекта H возможно использовать статистический подход.

4. Структура карты окружающего пространства

Иерархическая модель карты. На основе рассматриваемого подхода окружающее мобильного робота пространство представлено в виде определенным образом сгруппированных сегментов. С учетом относительного характера изменения координат (локализация объектов происходит в системе координат мобильного робота), для хранения данных удобно использовать древовидную структуру [5, 8, 9]. Корневым узлом структуры является точка, соответствующая началу отсчета в системе координат робота (центр масс робота). Каждому «листу» дерева соответствует полное или частичное отображение \mathcal{H}_p конкретного физического

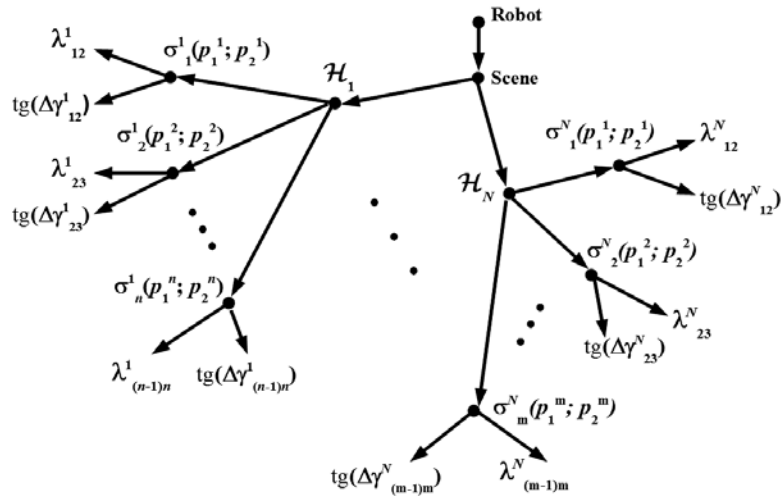


Рис. 1. Общий вид иерархической структуры для описания N объектов

объекта H_p . Наблюдаемым объектам также удобно сопоставить собственные локальные системы координат, что позволяет хранить их частные описания на нижеследующем уровне иерархической структуры. Общий вид рассматриваемой структуры показан на рис. 1.

Корневой узел, обозначенный на рисунке как Robot, и уровень наблюдаемой сцены Scene имеют абстрактное разделение, тогда как на физическом уровне эти точки совпадают. Хранение информации о количестве N выделенных объектов \mathcal{H}_i и вероятностных характеристиках их представлений на уровне сцены позволяет сократить необходимое количество операций поиска и сравнения для идентификации объектов или уточнения их моделей. Объекты $\mathcal{H}_1 \dots \mathcal{H}_N$ в соответствии с (3) представлены наборами сегментов $\sigma_j^i(p_1^j; p_2^j)$, где i – индекс объекта в описании сцены, j – индекс сегмента в его геометрической модели, пара точек $(p_1^j; p_2^j)$ соответствует координатам начала и конца j -го сегмента. Упорядочивание сегментов производится в соответствии с их расположением на контурах объекта таким образом, что разность $(p_2^{j-1} - p_1^j) \rightarrow \min$, и пара индексов $\{(j-1), j\}$ соответствует ближайшим сегментам. Для каждого из сегментов $\{\sigma_{j-1}^i; \sigma_j^i\}$ вычисляется отношение их длин $\lambda_{(j-1)j}^i$ и тангенс угла между ними $\text{tg}(\Delta\gamma_{(j-1)j}^i)$ в соответствии с выбранными в разделе 1 параметрами инвариантности геометрической модели.

ской модели.

Описание общей динамики системы. В рамках данной работы динамика робота и наблюдаемого пространства может быть описана следующими типами взаимных перемещений:

1. Робот перемещается среди статичных объектов, не изменяющих геометрическую конфигурацию. При этом наблюдателю известны точные или средние значения направления и величины перемещения, либо эти данные отсутствуют вообще (например, при включении системы).

2. Робот не перемещается, наблюдаемый мир может изменяться. Наблюдаемые объекты изменяют или свое положение в пространстве, и/или геометрическую конфигурацию.

3. Робот движется, объекты окружающего мира изменяются.

В общем случае движение наблюдателя (робота) в трехмерном пространстве может быть описано в терминах линейной алгебры как композиция вращения и сдвига. Вращение в евклидовом пространстве задается матрицей поворота:

$$\mathbf{R}(\theta_x, \theta_y, \theta_z) = \mathbf{M}_x(\theta_x)\mathbf{M}_y(\theta_y)\mathbf{M}_z(\theta_z),$$

где $\mathbf{M}_x(\theta_x)$, $\mathbf{M}_y(\theta_y)$, $\mathbf{M}_z(\theta_z)$ – матрицы поворота на угол θ вокруг осей X , Y и Z соответственно.

Обозначим \mathbf{d} вектор сдвига. Тогда движение наблюдателя может быть представ-

лено как

$$x' = \mathbf{R}(\theta_x, \theta_y, \theta_z) \cdot x + \mathbf{d}$$

или в матричной форме как

$$\begin{pmatrix} x' \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R}(\theta_x, \theta_y, \theta_z) & \mathbf{d} \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ 1 \end{pmatrix}. \quad (5)$$

В соответствии с представленной выше структурой карты, наблюдатель (робот) описан на корневом уровне древовидной структуры и соответствует нестационарной системе координат. Тогда движение наблюдателя может быть представлено как переход к новому базису. На нижележащих уровнях карты используются локальные системы координат со своими собственными матрицами перехода относительно движения корневого узла.

Пусть древовидная структура карты имеет две системы координат: корневую на уровне робота/сцены и локальную на уровне объектов \mathcal{H}_i (как показано на рис. 1). Обозначим \mathbf{P} матрицу перехода к локальной системе координат, в которой определяются координаты сегментов $\sigma_j^i(p_1^j; p_2^j)$ (рис. 1), в виде (5). Такая структура позволит изменять только матрицу перехода \mathbf{P} (применение процедуры перехода к новому базису), а сегменты, составляющие покрытие границ объекта со стационарной геометрической конфигурацией, будут инвариантны относительно движений наблюдателя.

Предположим, наблюдатель переместился, и матрица перехода \mathbf{T} имеет соответствующие компоненты вращения \mathbf{R} и

сдвига \mathbf{d} . Эта матрица может быть факторизована как

$$\mathbf{T} = \begin{pmatrix} \mathbf{R} & \mathbf{d} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{E}_3 & \mathbf{d} \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{R} & 0 \\ 0 & 1 \end{pmatrix},$$

где \mathbf{E}_3 — единичная матрица третьего порядка.

Тогда известная матрица перехода \mathbf{P} может быть преобразована как

$$\mathbf{P}' = \begin{pmatrix} \mathbf{R}^{-1} & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{E}_3 & -\mathbf{d} \\ 0 & 1 \end{pmatrix} \cdot \mathbf{P}.$$

Поскольку матрица поворота \mathbf{R} является ортогональной и ее определитель $\det \mathbf{R} = 1$, то ее обратная матрица \mathbf{R}^{-1} равна транспонированной \mathbf{R}^T . Такая факторизация позволяет упростить алгоритмы определения движения объектов (обнаружения перехода в связи с движением объекта) и изменения матриц перехода при собственном движении робота, что позволяет снизить вычислительные затраты.

Наблюдение динамических объектов. Как было сказано выше, следует учитывать, что наблюдаемые объекты не являются стационарными. В общем случае необходимо описать два вида их динамики: перемещение в пространстве и изменение геометрической конфигурации (формы). Изменение структуры карты для этих двух случаев схематически показано на рис. 2.

Пусть робот производит серию из N наблюдений через малые промежутки времени, т. е. работает в дифференциальном

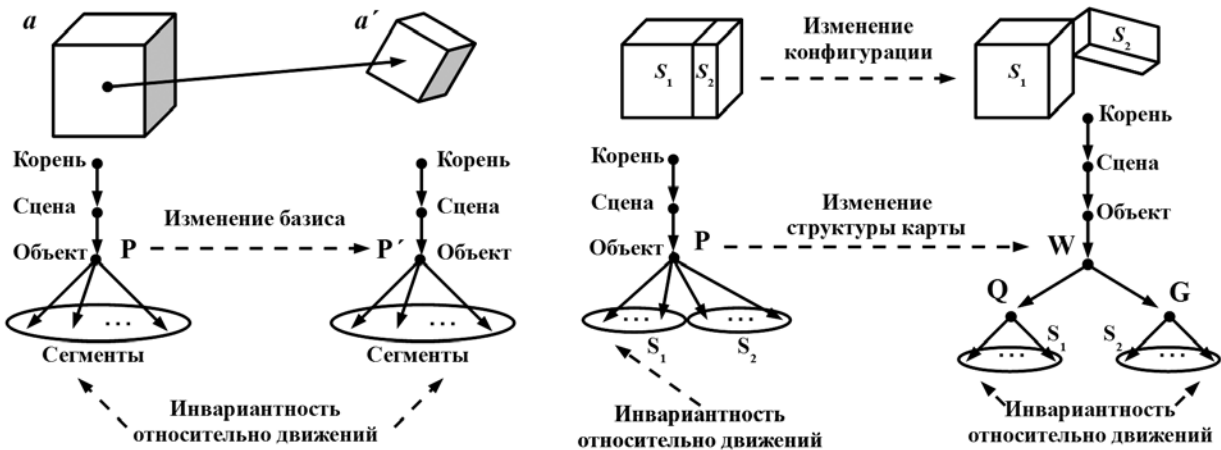


Рис. 2. Схематическое изображение изменения структуры карты для случаев перемещения объекта и его реконфигурации

режиме. Результаты представляют собой специализированный FIFO буфер, включающий $N+1$ моделей $\{\mu_r, \mu_{r-1}, \dots, \mu_{r-n}\}$, где μ_r соответствует текущему наблюдению, μ_{r-1} — предыдущему и т. д. В таком режиме возможно обнаружение последовательных изменений (тренда) на основе предположения, что общий тренд представляет собой композицию последовательных изменений.

Например, для набора $\{\mu_r, \mu_{r-1}, \mu_{r-2}\}$ $(\mu_r, \mu_{r-2}) = (\mu_r, \mu_{r-1}) \cdot (\mu_{r-1}, \mu_{r-2})$.

Движение объекта с сохранением формы (например, из положения a в положение a' на рис. 2) не изменяет инвариантных характеристик наблюдаемых моделей и затрагивает только положение нуля локальной системы координат. Это означает, что достаточно пересчитать только матрицу перехода \mathbf{P} к новому базису.

При изменении конфигурации инвариантное множество сегментов подвергается разбиению на подмножества в количестве $n \geq 2$ (S_1 и S_2 на рис. 2). При этом в границах данных подмножеств инвариантность сохраняется. Независимость каждого подмножества позволяет выделить для них собственные локальные системы координат. На рисунке показан переход к новым базисам с матрицами перехода \mathbf{Q} и \mathbf{G} . Учитывая, что подсистемы S_1 и S_2 связаны (присутствует исходная локальная система координат для объекта $S_1 \cup S_2$), необходимо ввести третью локальную систему координат с матрицей перехода \mathbf{W} . При этом изменение структуры карты происходит с увеличением глубины дерева на единицу.

Выделение изменений конфигурации объекта производится только в дифференциальном режиме на основе буферной модели наблюдения.

Идентификация объекта по его проекционной модели. На основе найденных величин и отношений можно предложить следующий метод идентификации наблюдаемых объектов:

- производится наблюдение объекта и построение проекционной модели μ' ;
- для проекционной модели вычисляются ее структурно-топологические характеристики: углы между отрезками и отношения сторон;

- производится поиск в дереве объекта, имеющего такой же набор топологических характеристик, что и проекционная модель μ' ;

- при нахождении заданного объекта (модель μ) вычисляется действие $g(\phi)$ и действие $g'(a)$;

- вычисленные действия g и g' применяются к проекционной модели μ' ;

- производится оценка равенства $\mu \cap \mu' = \mu'$.

В случае равенства можно утверждать, что наблюдаемый объект ранее описан и имеет геометрическую модель μ , а наблюдатель находится в такой позиции относительно объекта, которая соответствует вычисленным действиям $g(\phi)$ и $g'(a)$.

Этот подход дает однозначное решение задачи в случае, если модель μ не имеет двух и более пар отрезков с эквивалентными геометрическими характеристиками. Множество из двух и более эквивалентных пар приведет к тому, что описанный выше метод даст множество пар действий $\{g, g'\}$. В этом случае метод позволяет подтвердить, что наблюдаемые части имеются в составе модели объекта, однако данных для идентификации наблюдаемой части (из множества возможных эквивалентных частей) и положения относительно объекта недостаточно.

Рассчитаем необходимый объем вычислительных операций для приведенного алгоритма. Пусть проекционная модель μ' содержит M сегментов. Для данной модели необходимо произвести следующие вычисления:

длин (или квадратов длин) отрезков — M операций;

тангенсов углов наклона (к осям системы координат в трехмерном пространстве) — $2M$ операций;

тангенсов углов между отрезками — $M(M-1)$ операций.

Таким образом, всего требуется $M^2 + 2M$ вычислительных операций.

Рассмотрим случай прямого неоптимизированного поиска объекта, описываемого моделью μ' , в дереве. Пусть дерево содержит в общей сложности T отрезков. Тогда процедура поиска будет состоять из сравнения

тангенсов углов между отрезками модели μ' с тангенсами углов между отрезками, храняемыми в дереве. Всего получаем операций сравнения

$$\frac{T(T-1)}{2} \frac{M(M-1)}{2}$$

Очевидно, что при увеличении T время процедуры поиска по дереву будет возрастать квадратично при условии того, что среднее количество отрезков, наблюдаемых в проекционной модели μ' , не превышает некоторого M_{\max} . Поэтому целесообразно применить дифференциальный подход, использующий данные о проекционной модели на предыдущем этапе наблюдения. Как было замечено выше, применить такой подход возможно при незначительных изменениях в наблюдаемых проекционных моделях, что достигается за счет повышения частоты процедур наблюдения.

Пусть имеем проекционную модель μ' , описывающую предыдущее наблюдение, и проекционную модель μ'' , описывающую обрабатываемое наблюдение. Пусть количество отрезков модели $\mu' - M'$, в модели $\mu'' - M''$. Можно предполагать, что M'

сравнимо с M'' , то есть $M' \approx M'' \approx M$. Тогда приблизительное число операций сравнения составит

$$\frac{M^2(M-1)^2}{4} \leq \frac{M_{\max}^2(M_{\max}-1)^2}{4}$$

Например, при ограничении сверху $M_{\max} = 500$ получаем приблизительно $15,625 \cdot 10^9$ операций сравнения.

Дополнительные сегменты, которые составляют разницу между моделями μ' и μ'' , добавляются в итоговую структуру представления объектов.

Для уменьшения количества операций сравнения можно воспользоваться свойством малых отклонений модели μ'' от модели μ' . В случае совпадения пар отрезков двух моделей в дальнейшем они не участвуют в процедуре покомпонентного сравнения μ' и μ'' , что значительно снижает количество операций сравнения. Например, если модель μ' содержит $N' = 3$ отрезка, а модель $\mu'' - N'' = 4$ отрезка, и только первая процедура сравнения дала положительный итог, результирующее количество операций сравнения снижается на 38,89 % (11 операций вместо $3(3-1)4(4-1)/4 = 18$).

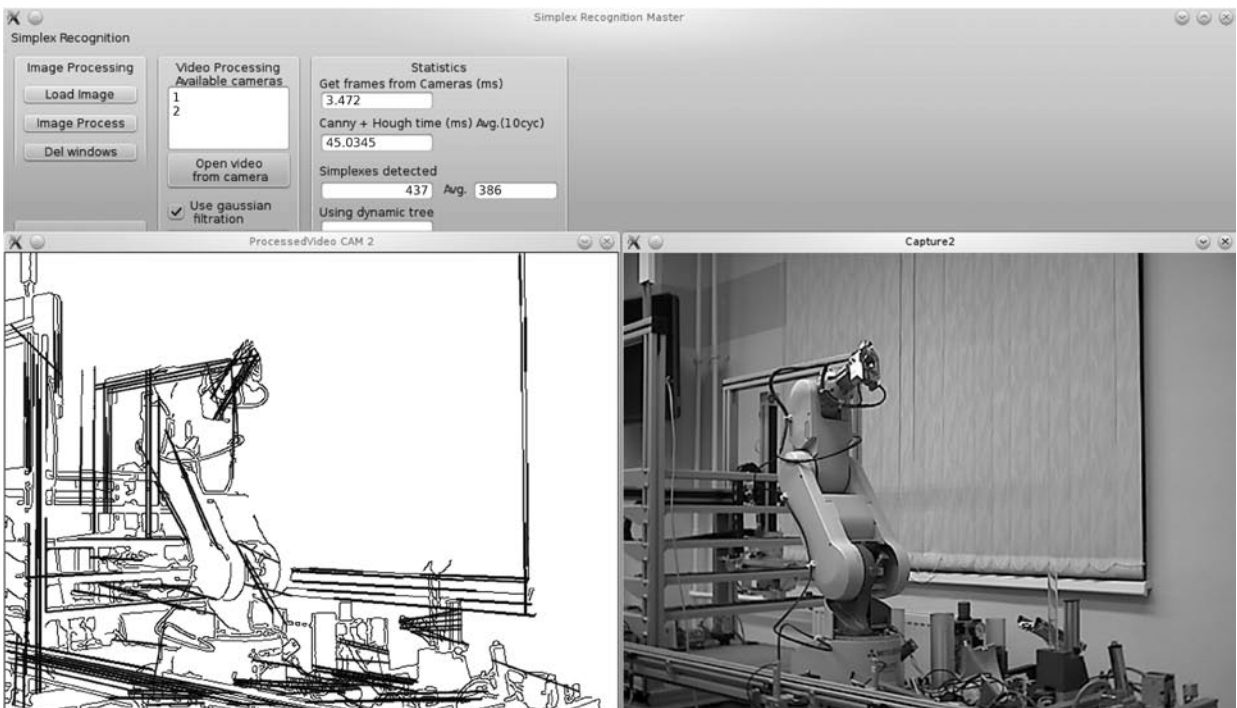


Рис. 3. Пример работы экспериментального программного комплекса

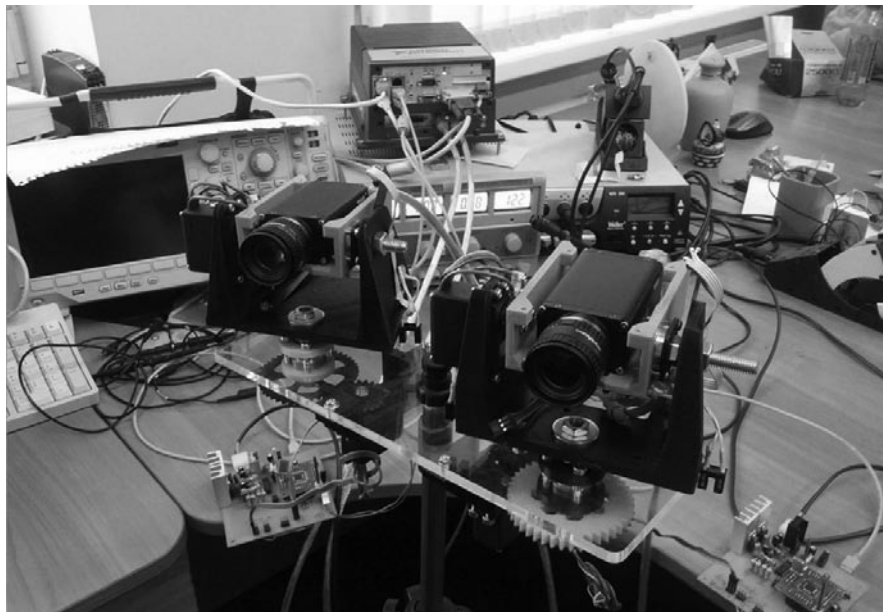


Рис. 4. Прототип мобильной автономной роботизированной системы

5. Экспериментальные результаты

Изначально подход к геометрическому описанию пространства на основе гранично-ориентированных проекций его образов был проверен с использованием библиотеки OpenCV на тестовом стенде со следующей конфигурацией: CPU – Intel Core2Duo 2 ГГц; RAM – DDR II 2 ГБ; стереопара – две цифровые камеры с разрешением $640 \times 480 \times 24$ бит и шиной USB2.0; программное обеспечение – OS Debian GNU/Linux, ker.3.0, KDE4, Qt4, OpenCV2.1 [8, 9].

В результате были получены исходные данные для построения моделей объектов с различными структурами наблюдаемых поверхностей и при разных условиях наблюдения. Пример работы тестового программного комплекса для описания геометрических объектов сложной формы показан на рис. 3. Для определения границ использовался детектор Кенни (Canny) [14]. Выделение аналитически заданных примитивов осуществлялось с использованием преобразования Хафа (Hough Transform) [15]. Эти методы подробно описаны и широко применяются для решения различных прикладных задач, что обусловило их использование для предварительной оценки

скорости работы и особенностей реализации разрабатываемого подхода.

В ходе проверки на тестовом стенде было установлено, что суммарное время обработки потока данных со стереоскопической пары не превышало 50 мс, а уровень загрузки центрального процессора не поднимался выше 38 %.

Положительные результаты позволили приступить к разработке аппаратного комплекса на основе встраиваемой системы компьютерного зрения EVS-1464RT производства National Instruments с подключением двух цифровых камер высокого разрешения Basler Scout sA1390-17fc. Общий вид аппаратного комплекса показан на рис. 4.

Были разработаны оригинальные электронные и механические подсистемы управления положением камер и фокусировки объективов, приложение для настройки и обработки изображения с распределенной клиент-серверной архитектурой [16]. В настоящее время программное обеспечение проходит доработку с целью внедрения более совершенного по сравнению с алгоритмом Кенни метода определения границ на основе представления изображения в частотной области [10], а также оптимизации структуры итоговой динамической карты.

Данная статья развивает предложенный в работах [8, 9] метод полного и инвариантного представления окружающего пространства для мобильных автономных систем, действующих в стохастической, динамически изменяющейся среде в условиях отсутствия предварительных сведений о ней. Метод построения динамических геометрических моделей объектов окружающего пространства (динамических карт)

может служить основой для разработки алгоритмов автономного целенаправленного перемещения мобильных роботов. Для реализации описанных методов нами разрабатывается прототип мобильной автономной роботизированной системы для решения таких прикладных задач как, например, проведение разведки, поиска и перемещения объектов в заданную точку при отсутствии исходных данных о локации.

СПИСОК ЛИТЕРАТУРЫ

1. **Yasuyuki S.** Japan's Decline as a Robotics Superpower: Lessons From Fukushima // *The Asia-Pacific Journal*. 2011. Vol. 9. Iss. 24. No. 2.
2. **Guizzo E.** Robot Vacuum Sucks Up Radiation at Fukushima Plant // *IEEE Spectrum*. 07.07.2011 [электронный ресурс] / URL: <http://spectrum.ieee.org/automaton/robotics/industrial-robots/robot-vacuum-sucks-up-radiation-at-fukushima> (дата обращения: 03.06. 2015).
3. **Guizzo E.** Fukushima Robot Operator Writes Tell-All Blog // *IEEE Spectrum*. 23.08.2011 [электронный ресурс] / URL: <http://spectrum.ieee.org/automaton/robotics/industrial-robots/fukushima-robot-operator-diaries> (дата обращения: 03.06.2015).
4. **Siciliano B., Khatib O.** (eds.). *Springer Handbook of Robotics*. Springer, 2008.
5. **Арсеньев Д.Г., Бабакина Н.А., Колесников М.П., Шкодырев В.П.** Построение векторных моделей окружающего пространства для задач управления в мобильных робототехнических системах // *Робототехника и техническая кибернетика*. 2014. № 1(2). С. 58–64.
6. **Nuchter A.** 3D Robotic Mapping: The Simultaneous Localization and Mapping Problem with Six Degrees of Freedom. Springer, 2009.
7. **Визильтер Ю.В., Желтов С.Ю., Бондаренко А.В. и др.** Обработка и анализ изображений в задачах машинного зрения. М.: Физматкнига, 2010. 672 с.
8. **Бабакина Н.А., Колесников М.П.** Построение динамических геометрических моделей окружающего пространства для мобильных автономных систем // *Информационно-управляющие*

- системы. 2012. № 1(56). С. 51–59.
9. **Babakina N.A., Kolesnikov M.P.** Using of environment's dynamic geometric models for solving the mobile robot SLAM problem // *Proc. of 5th Internat. Symp. on Communications Control and Signal Processing*. Roma, 2012.
10. **Kovesi P.** Phase Congruency Detects Corners and Edges // *The Australian Pattern Recognition Society Conference: DICTA*. Sydney, 2003. Pp. 309–318.
11. **Рурк К., Сандерсон Б.** Введение в кусочно-линейную топологию. М.: 1974. 208 с.
12. **Пергат А.** Нечеткое моделирование и управление. М.: БИНОМ. Лаборатория знаний, 2009. 798 с.
13. **Ширяев А.Н.** Вероятность: в 2 кн. Кн. 1. 3-е изд., перераб. и доп. М.: МЦНМО, 2004. 520 с.
14. **Canny J.** A computational approach to edge detection // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1986. Vol. PAMI-8. No. 6. Pp. 679–698.
15. **Ballard D.H.** Generalizing the Hough transform to detect arbitrary shapes // *Pattern Recognition*. 1981. Vol. 13. No. 2. Pp. 111–122.
16. **Арсеньев Д.Г., Бабакина Н.А., Колесников М.П., Шкодырев В.П.** Построение векторных моделей окружающего пространства для задач управления в мобильных робототехнических системах // *Труды VII междунар. симп. Экстремальная робототехника – робототехника для работы в условиях опасной окружающей среды*. СПб., 2013. С. 471–478.

REFERENCES

1. **Yasuyuki S.** Japan's Decline as a Robotics Superpower: Lessons From Fukushima. *The Asia-Pacific Journal*, 2011, Vol. 9, Issue 24, No. 2.
2. **Guizzo E.** *Robot Vacuum Sucks Up Radiation*

- at Fukushima Plant, IEEE Spectrum*, July 07, 2011. Available: <http://spectrum.ieee.org/automaton/robotics/industrial-robots/robot-vacuum-sucks-up-radiation-at-fukushima> (Accessed: 03.06. 2015).

3. **Guizzo E.** *Fukushima Robot Operator Writes Tell-All Blog*, *IEEE Spectrum*, August 23, 2011. Available: <http://spectrum.ieee.org/automaton/robotics/industrial-robots/fukushima-robot-operator-diaries> (Accessed: 03.06.2015).
4. **Siciliano B., Khatib O.** (eds.) *Springer Handbook of Robotics*. Springer, 2008.
5. **Arsenyev D.G., Babakina N.A., Kolesnikov M.P., Shkodyrev V.P.** Postroyeniye vektornykh modeley okruzhayushchego prostranstva dlya zadach upravleniya v mobilnykh robototekhnicheskikh sistemakh [Design of vector-based environment models for the control problems in mobile robotic systems]. *Robototekhnika i tekhnicheskaya kibernetika* [Robotics and Technical Cybernetics], 2014, No. 1(2), Pp. 58–64. (rus)
6. **Nuchter A.** *3D Robotic Mapping: The Simultaneous Localization and Mapping Problem with Six Degrees of Freedom*. Springer, 2009.
7. **Vizilter Yu.V., Zheltov S.Yu., Bondarenko A.V. et al.** *Obrabotka i analiz izobrazheniy v zadachakh mashinnogo zreniya* [Image processing and analysis in machine vision issues]. Moscow: Fizmatkniga Publ., 2010, 672 p. (rus)
8. **Babakina N.A., Kolesnikov M.P.** Postroyeniye dinamicheskikh geometricheskikh modeley okruzhayushchego prostranstva dlya mobilnykh avtonomnykh sistem [Construction of Dynamic Geometric Models of the Environment for Mobile Autonomous Systems]. *Informatsionno-upravlyayuschie sistemy* [Information and Control Systems], 2012, No. 1(56), Pp. 51–59. (rus)
9. **Babakina N.A., Kolesnikov M.P.** Using of environment's dynamic geometric models for solving the mobile robot SLAM problem. *Proceedings of 5th International Symposium on Communications Control and Signal Processing*. Roma, 2012.
10. **Kovesi P.** Phase Congruency Detects Corners and Edges. *The Australian Pattern Recognition Society Conference: DICTA*. Sydney, 2003. Pp. 309–318.
11. **Rurk K., Sanderson B.** *Vvedeniye v kusochno lineynuyu topologiyu* [Introduction to piecewise-linear topology]. Moscow, 1974, 208 p. (rus)
12. **Pegat A.** *Nechetkoye modelirovaniye i upravleniye* [Fuzzy modeling and control]. Moscow: BINOM. Laboratoriya znaniy Publ., 2009, 798 p. (rus)
13. **Shiryayev A.N.** *Veroyatnost: v 2 kn. Kniga 1.* [Probability: in 2 books. Book 1]. Moscow: MCNMO Publ., 2004, 520 p. (rus)
14. **Canny J.** A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, Vol. PAMI-8, No. 6, Pp. 679–698.
15. **Ballard D.H.** Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 1981, Vol. 13, No. 2, Pp. 111–122.
16. **Arsenyev D.G., Babakina N.A., Kolesnikov M.P., Shkodyrev V.P.** Postroyeniye vektornykh modeley okruzhayushchego prostranstva dlya zadach upravleniya v mobilnykh robototekhnicheskikh sistemakh [Design of vector-based environment models for the control problems in mobile robotic systems]. *Trudy VII mezhdunarodnogo simpoziuma Ekstremalnaya robototekhnika – robototekhnika dlya raboty v usloviyah opasnoy okruzhayushey sredy* [Proceedings of 7th IARP RISE-ER'2013]. St. Petersburg, 2013, Pp. 471–478. (rus)

БАБАКИНА Надежда Александровна – старший преподаватель кафедры систем и технологий управления Института компьютерных наук и технологий Санкт-Петербургского политехнического университета Петра Великого.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: nadia.babakina@gmail.com

БАБАКИНА Nadezhda A. Peter the Great St. Petersburg Polytechnic University.

195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.

E-mail: nadia.babakina@gmail.com

КОЛЕСНИКОВ Максим Павлович – инженер 1 категории ОАО «КОНЦЕРН «ЦНИИ «ЭЛЕКТРОПРИБОР».

197046, Россия, Санкт-Петербург, улица Малая Посадская, д. 30.

E-mail: maxim.p.kolesnikov@gmail.com

KOLESNIKOV Maxim P. Concern CSRI Elektropribor, JSC.

197046, Malaya Posadskaya Str. 30, St. Petersburg, Russia.

E-mail: maxim.p.kolesnikov@gmail.com

ШКОДЫРЕВ Вячеслав Петрович – директор научно-исследовательского комплекса «Математическое моделирование и системы управления» Санкт-Петербургского политехнического университета Петра Великого, доктор технических наук.

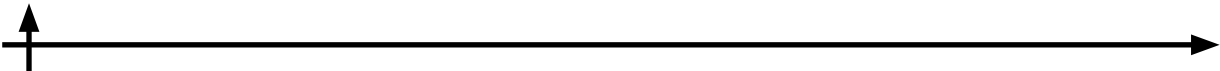
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: shkodyrev@mail.ru

SHKODYREV Viacheslav P. *Peter the Great St. Petersburg Polytechnic University.*

195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.

E-mail: shkodyrev@mail.ru



Программное обеспечение вычислительных, телекоммуникационных и управляющих систем

DOI: 10.5862/JCSTCS.229.3

УДК 681.7.055.4

Е.А. Никулина

РАСЧЕТ ФУНКЦИИ РАССЕЯНИЯ ТОЧКИ С УЧЕТОМ ДВУЛУЧЕПРЕЛОМЛЕНИЯ И ЧАСТИЧНОЙ КОГЕРЕНТНОСТИ ОСВЕЩЕНИЯ С ИСПОЛЬЗОВАНИЕМ ГРАФИЧЕСКИХ ПРОЦЕССОРОВ

Е.А. Nikulina

POINT-SPREAD FUNCTION COMPUTATION UNDER THE CONDITIONS OF BIREFRINGENCE AND PARTIAL COHERENCE OF LIGHTING USING GPUS

Рассмотрено применение метода представления комплексной амплитуды через суперпозицию плоских волн для учета влияния двулучепреломления в фотолитографических системах. Изображение в проходящем свете при частично когерентном освещении проводится по методу интегрирования по источнику. Показана вычислительная сложность предложенной методики, а также отмечены преимущества использования графических процессоров для решения такого рода задач. Описаны особенности использования графических процессоров для решения неграфических задач. Приведены оценки временных затрат при расчете по описанной методике с применением как графических, так и традиционных процессоров. Показаны результаты моделирования функции рассеяния точки в различных условиях.

ДВУЛУЧЕПРЕЛОМЛЕНИЕ; ПОЛЯРИЗАЦИЯ; ФУНКЦИЯ РАССЕЯНИЯ ТОЧКИ; ПАРАЛЛЕЛЬНЫЕ ВЫЧИСЛЕНИЯ; CUDA.

This paper considers an application of the method of representing a complex amplitude as a superposition of plane waves to take birefringence in photolithographic systems into account. The image in transmitted light under partially coherent lighting is conducted according to the source integral method. This paper shows the computational costs of the proposed method, also mentioning the benefits of GPU computation for problem-solving of this type. Special aspects of GPU computation for non-graphical problem-solving are described. This paper introduces estimates of span time in case of computation by means of the mentioned method with GPU as well as CPU. This paper shows the results of simulating the point-spread function under different conditions.

BIREFRINGENCE; POLARIZATION; POINT SPREAD FUNCTION; PARALLEL COMPUTING; CUDA.

Несмотря на широкое распространение и использование во многих сферах научной и повседневной жизни полупроводниковых микросхем, их производство остается одним из сложнейших технологических процессов за всю историю человечества.

Как известно, основным способом производства полупроводниковых микросхем является фотолитография, включающая в себя сложные технологические операции, в том числе отжиг стеклянной заготовки, который сокращает остаточные неравномер-

ности материала. Однако даже при весьма длительном отжиге стекла в нем наблюдаются остаточные напряжения, в результате чего в оптически изотропном от природы стекле возникает оптическая анизотропия, которая проявляется в двойном лучепреломлении. При падении световой волны на анизотропную среду в ней возникают две волны с взаимно перпендикулярными плоскостями поляризации, скорость распространения одной волны не зависит от направления распространения (обыкновенный луч), а скорость распространения другой волны зависит от направления распространения (необыкновенный луч). Кроме того, к остаточным напряжениям, определяющим появление двойного лучепреломления, приводят структурные изменения при механической обработке стекла и кристаллических материалов. В оптических системах для микролитографии наличие остаточного двулучепреломления в материалах линз может стать причиной заметного ухудшения качества изображения [1].

Ухудшение качества изображения обусловлено влиянием на разрешающую способность изображения деформации волнового фронта, выраженным в длинах волн применяемого излучения. Таким образом, влияние деформации волнового фронта, вызванного двулучепреломлением, может стать очень серьезным, если используется излучение с длиной волны, равной 193 нм, или с длиной волны, равной 157 нм [2, 3].

Для определения предельного разрешения изображения необходимо учитывать недостаток скалярных представлений при вычислении распределения освещенности в дифракционном изображении точки, а также частичную когерентность источников излучения [4]. Все это требует разработки новых алгоритмов расчета формирования изображения, основанных на более точном применении дифракционной теории в случае присутствия двулучепреломления. Также, в связи с невозможностью отбраковывания по величине двулучепреломления заготовок, возникает необходимость учета и компенсации двулучепреломления в оптических высокоапертурных фотолитографических системах.

В связи с многократным расчетом передаточной функции применение обычных скалярных методов расчета функции рассеяния точки в случае частично когерентного освещения приводит к значительному увеличению вычислительных нагрузок по сравнению со случаем полностью когерентного освещения.

Более строгое применение дифракционной теории, с учетом остаточного двулучепреломления в материалах линз, также приводит к дополнительному приросту вычислений. Кроме того, высочайшие требования к качеству изображения фотолитографических систем приводят к увеличению дискретизации, а следовательно, и к увеличению объемов обрабатываемых данных. Таким образом, кроме оптимизации самих методик расчета также необходимо обратить пристальное внимание на программную реализацию расчетных алгоритмов.

Для применения строгой теории дифракции в случае высокоапертурных систем используют представление комплексной амплитуды монохроматического поля через суперпозицию базисных плоских волн.

Базисная плоская волна [5] имеет единичную амплитуду с линейной поляризацией, направление ее распространения в пространстве определяется направляющими косинусами c_x , c_y и c_z . Состояние поляризации базисной плоской волны может изменяться при взаимодействии с поляризирующими средами.

Плоские волны, которые удовлетворяют перечисленным условиям, будем считать элементами базиса аппроксимации комплексной амплитуды монохроматического поля произвольного вида, причем комплексные коэффициенты аппроксимации будут представлять собой спектр векторных плоских волн.

Комплексную амплитуду монохроматического поля можно представить в виде суперпозиции базисных векторных плоских волн единичной амплитуды:

$$\mathbf{E} = \sum_i \sum_j \begin{pmatrix} \bar{f}_{ij}^x & 0 & 0 \\ 0 & \bar{f}_{ij}^y & 0 \\ 0 & 0 & \bar{f}_{ij}^z \end{pmatrix} \cdot \mathbf{E}_{ij}, \quad (1)$$

где \bar{f}_{ij}^x , \bar{f}_{ij}^y и \bar{f}_{ij}^z – коэффициенты разложения поля в спектр плоских волн;

$\mathbf{E}_{ij} = \frac{\mathbf{k}_{ij} \times \mathbf{p}_{ij}}{|\mathbf{k}_{ij} \times \mathbf{p}_{ij}|}$ – базисная векторная плоская волна единичной амплитуды; \mathbf{p}_{ij} – вектор поляризации, определяющий состояние поляризации, перпендикулярный векторной плоской волне \mathbf{E}_{ij} ;

$$\mathbf{k} = \frac{2\pi}{\lambda} \begin{pmatrix} c_x \\ c_y \\ c_z \end{pmatrix} = \frac{2\pi}{\lambda} \begin{pmatrix} A'_0 \rho_x \\ A'_0 \rho_y \\ \sqrt{1 - (A'_0)^2 (\rho_x^2 + \rho_y^2)} \end{pmatrix} \text{ – вол-}$$

новой вектор, записанный в канонических координатах, направление плоской волны

$$\mathbf{R}_z = \begin{pmatrix} \cos^2 \theta + \sin^2 \theta e^{-i\delta} & \cos \theta \cdot \sin \theta (1 - e^{-i\delta}) & 0 \\ \cos \theta \cdot \sin \theta (1 - e^{-i\delta}) & \sin^2 \theta + \cos^2 \theta e^{-i\delta} & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ –}$$

матрица направления поляризации; θ , δ –

угол и величина поляризации; $\mathbf{V}_m = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ –

вектор Максвелла–Джонса в данном случае описывает линейную поляризацию.

Перемножив векторы, в соответствии с формулой 2 нетрудно найти коэффициенты разложения комплексной амплитуды поля по спектру плоских волн. Далее распространение поля через оптическую систему необходимо рассматривать отдельно для трех компонент комплексной амплитуды. Если не учитывать дифракционное рассеяние на элементах фазосдвигающего шаблона, то компоненты комплексной амплитуды поля можно считать равными комплексной амплитуде поля на входном зрачке, вычисленной так же, как и при использовании скалярной модели.

Моделирование формирования изображения при частично-когерентном освещении производится с использованием метода интегрирования по источнику [7, 8].

Источник представляется как совокупность точек, каждая из которых создает полностью когерентное освещение предмета, а оптическая система – соответствующее ему изображение. Описание распространения электромагнитного поля от предмета к изо-

задано при помощи направляющих косинусов c_x , c_y и c_z , A'_0 – задняя апертура проекционного объектива, ρ_x и ρ_y – канонические координаты на зрачке.

После прохождения двулучепреломляющего материала обыкновенные и необыкновенные лучи, распространяющиеся в одном направлении, складываются таким образом, что результирующий луч становится эллиптически поляризованным [6]. Следовательно, вектор поляризации можно вычислить по формуле:

$$\mathbf{p}_{ij} = \mathbf{k}_{ij} \times (\mathbf{R}_z \cdot \mathbf{V}_m), \quad (2)$$

где

бражению осуществляется с использованием дифракционной теории:

$$\mathbf{U}_4 = \begin{pmatrix} u_4^x \\ u_4^y \\ u_4^z \end{pmatrix} =$$

$$\begin{pmatrix} F^{-1}\{F\{U_0(\eta_x, \eta_y) \cdot T(\eta_x, \eta_y)\} \cdot f \cdot \bar{f}^x\} \\ F^{-1}\{F\{U_0(\eta_x, \eta_y) \cdot T(\eta_x, \eta_y)\} \cdot f \cdot \bar{f}^y\} \\ F^{-1}\{F\{U_0(\eta_x, \eta_y) \cdot T(\eta_x, \eta_y)\} \cdot f \cdot \bar{f}^z\} \end{pmatrix}, \quad (3)$$

где \mathbf{U}_4 – распределение комплексной амплитуды поля на изображении; \bar{f}^x , \bar{f}^y и \bar{f}^z – коэффициенты разложения комплексной амплитуды поля по спектру плоских волн; f – зрачковая функция; $U_0(\eta_x, \eta_y) = \exp[2\pi i(\eta_x \xi_x^S + \eta_y \xi_y^S)]$ – комплексная амплитуда поля на предмете от отдельной точки источника; $T(\eta_x, \eta_y)$ – функция, описывающая форму, размеры и комплексное пропускание предмета.

Распределение интенсивности изображения, формируемого отдельной точкой источника, определяется следующим выражением:

$$I'(\eta'_x, \eta'_y) = |u_4^x|^2 + |u_4^y|^2 + |u_4^z|^2. \quad (4)$$

Таким образом, для определения распределения интенсивности используются все три декартовых составляющих вектора

поля. Итоговое распределение интенсивности изображения, полученного от источника произвольной формы, вычисляется интегрированием изображений, полученных от каждой точки источника, по интенсивности.

Предварительные результаты исследования влияния двулучепреломления по описанной методике в реальных образцах флюорита рассмотрены в статье [9].

Применение плоских волн для математического моделирования полей удобно тем, что можно так же, как и в модели на основе скалярной теории, использовать аппарат преобразования Фурье. Несмотря на относительно небольшое усложнение расчетов по сравнению со скалярной моделью, расчет по данной методике займет приблизительно в три раза больше времени. Для оценки объема требуемых вычислений заметим, что размер выборки обычно составляет 1024×1024 точки, степень когерентности источника достигает 0,5, и для каждой его точки требуется выполнить шесть преобразований Фурье по всему полю.

Проблема объемных потоковых вычислений постоянно возникает во многих современных областях науки. Для увеличения скорости потоковых вычислений и обработки большого объема данных в последнее время получило активное развитие новое направление – использование графических процессоров для решения неграфических задач.

Одним из флагманов этого движения является технология CUDA – программно-аппаратная вычислительная архитектура NVIDIA, основанная на расширении языка C, которая дает возможность организации доступа к набору инструкций графического ускорителя (GPU) и управления его памятью при организации параллельных вычислений.

Основными преимуществами технологии CUDA являются ее простота – все программы пишутся на «расширенном» языке C, наличие хорошей документации, набор готовых инструментов, набор готовых библиотек (например, библиотека, позволяющая выполнять быстрое преобразование Фурье для больших массивов), кроссплатформенность [10].

Для решения задач CUDA использует очень большое количество параллельно выполняемых нитей (потоков), при этом

обычно каждой нити соответствует один элемент вычисляемых данных. За счет подобного распараллеливания и появляется многократный прирост производительности относительно обычных вычислений при помощи центрального процессора (CPU) компьютера, выполняющего расчет всех нитей последовательно.

Идеология использования CUDA строится на концепции о том, что GPU выступает в роли массивно-параллельного сопроцессора к CPU. Программа на CUDA задействует как CPU, так и GPU. При этом обычный (последовательный, т. е. непараллельный) код выполняется на CPU, а для массивно-параллельных вычислений соответствующий код выполняется на GPU как набор одновременно выполняющихся нитей [11].

Использование технологии CUDA позволило увеличить производительность вычислений в десятки раз. Для расчета выборки размером 256×256 по разработанному методу базисных плоских волн с фактором когерентности 0,5 при помощи обычного процессора потребовалось 11 мин 12,4 с, а с использованием технологии CUDA – 0,1 с. Полученная в этом случае функция рассеяния точки приведена на рис. 1. Расчет выборки 1024×1024 при тех же параметрах при помощи технологии CUDA занял всего 10,9 с.

Для оценки влияния двулучепреломления на качество изображения проекционных

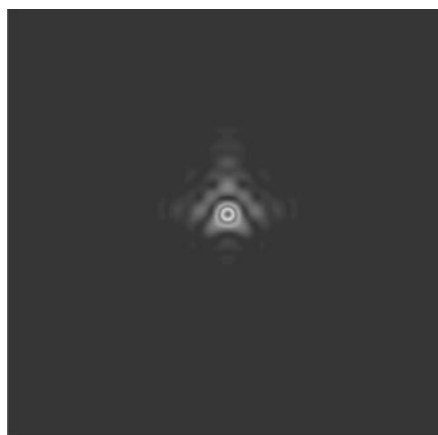


Рис. 1. ФРТ (центральный максимум 0,6) при частично когерентном освещении (степень когерентности 0,5) в присутствии двулучепреломления (0–8 нм/см) с учетом среднеквадратического отклонения волнового фронта 0,098 длины волны

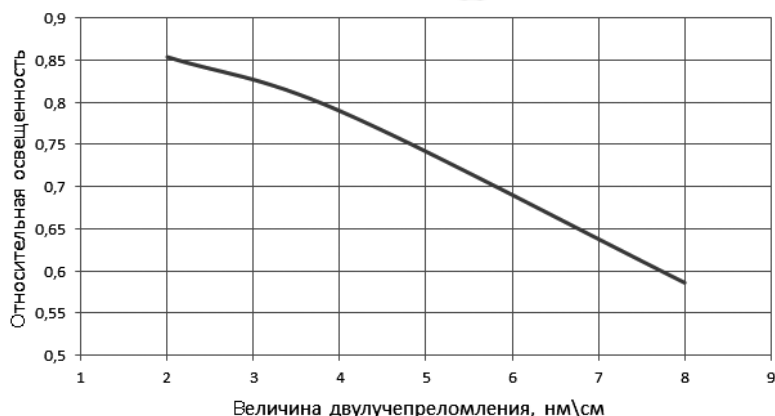


Рис. 2. График зависимости величины относительной освещенности в центре кружка Эри от средней величины двулучепреломления
Дополнительное среднеквадратическое отклонение волнового фронта 0,098 длины волны

фотолитографических систем проведено моделирование подобных систем при разных величинах двулучепреломления, с учетом волновых aberrаций этих систем и частичной когерентности источника. Коэффициент когерентности принимался равным 0,3.

На рис. 2 показан график зависимости величины относительной освещенности в центре кружка Эри от средней величины двулучепреломления. Величина освещенности оценивается относительно освещенности в центре дифракционного изображения точки оптической системой с такими же физическими характеристиками, но при отсутствии aberrаций и любых других искажающих факторов. Таким образом, относительная освещенность, равная единице, соответствует идеальной системе.

Приведенные исследования показали, что при разработке фотолитографических систем крайне важно учитывать не только волновые aberrации, но и остаточное двулучепреломление оптических элементов. Величина перераспределения энергии определяет относительную потерю освещенности в центре дифракционного пятна, а ее расчет можно произвести по описанной методике.

Использование технологии параллельных вычислений на графическом процессоре позволяет в этом случае значительно уменьшить время расчета. Главная особенность использования графических процессоров для решения вычислительных задач — необходимость постановки одинакового условия (например, математической формулы) для объемного массива данных, что позволяет эффективно обрабатывать параллельные потоки вычислений. Такой подход требуется для решения многих оптических задач, при которых очень часто возникает необходимость обработки множества лучей по одинаковым законам, но с разными входными данными.

Благодаря встроенной реализации алгоритмов быстрого преобразования Фурье предложенная методика расчета функции рассеяния точки легко поддавалась оптимизации по концепции параллельных вычислений. Однако графические процессоры могут использоваться и для решения многих других физических задач.

СПИСОК ЛИТЕРАТУРЫ

1. Hlubina P., Ciprian D. Absolute phase birefringence dispersion in polarization-maintaining fiber or birefringent crystal retrieved from a channeled spectrum // Optics Letters. 2010. Vol. 35. No. 10. Pp. 1566–1568.
2. Sarani A., Abdulhalim I. Spectropolarimetric method for optical axis, retardation, and birefringence dispersion measurement // Opt. Eng. 2009. Vol. 48.

- No. 5. P. 053601.
3. Sang-Kon Kim Polarized Effects in Optical Lithography with High NA Technology // Journal of the Korean Physical Society. 2007. Vol. 50. No. 6. Pp. 1952–1958.
4. Yanqiu Li, Xuejia Guo, Xiaolin Liu, Lihui Liu A technique for extracting and analyzing the polarization aberration of hyper-numerical aperture

image optics // Proc. SPIE 9042, Internat. Conf. on Optical Instruments and Technology: Optical Systems and Modern Optoelectronic Instruments. 2013.

5. **Домненко В.М., Бурсов М.В., Иванова Т.В.** Моделирование формирования оптического изображения: Учеб. пособие. СПб.: НИУ ИТМО, 2011. 141 с.

6. **Ландсберг Г.С.** Оптика: Учеб. пособие для вузов. 6-е изд., стереот. М.: Физматлит, 2003. 848 с.

7. **Иванова Т.В., Зуева Л.В.** Исследование способов дискретизации источника при моделировании фотолитографического изображения // Оптический журнал. 2012. Т. 79. № 5. С. 48–52.

8. **Можаров Г.А.** Теория аберраций оптических систем. СПб.: Изд-во «Лань», 2013. 288 с.

9. **Никулина Е.А., Зверев В.А.** Исследование влияния двулучепреломления на качество изображения фотолитографических систем при частично когерентном освещении // Научно-технический вестник информационных технологий, механики и оптики. 2015. Т. 15. № 2. С. 196–201.

10. Введение в технологию CUDA – Компьютерная графика и мультимедиа [электронный ресурс]// URL: <http://cgm.computergraphics.ru/issues/issue16/cuda> (дата обращения: 28.05.2015).

11. **Боресков А.В., Харламов А.А.** Основы работы с технологией CUDA. 2010. С. 20–33.

REFERENCES

1. **Hlubina P., Ciprian D.** Absolute phase birefringence dispersion in polarization-maintaining fiber or birefringent crystal retrieved from a channeled spectrum. *Optics Letters*, 2010, Vol. 35, No. 10, Pp. 1566–1568.

2. **Sarani A., Abdulhalim I.** Spectropolarimetric method for optical axis, retardation, and birefringence dispersion measurement. *Opt. Eng.*, 2009, Vol. 48, No. 5, P. 053601.

3. **Sang-Kon Kim.** Polarized Effects in Optical Lithography with High NA Technology. *Journal of the Korean Physical Society*, 2007, Vol. 50, No. 6, Pp. 1952–1958.

4. **Yanqiu Li, Xuejia Guo, Xiaolin Liu, Lihui Liu.** A technique for extracting and analyzing the polarization aberration of hyper-numerical aperture image optics. *Proc. SPIE 9042, International Conference on Optical Instruments and Technology: Optical Systems and Modern Optoelectronic Instruments*, 2013.

5. **Domnenko V.M., Bursov M.V., Ivanova T.V.** Modelirovaniye formirovaniya opticheskogo izobrazheniya [Simulation of optical imaging]. St. Petersburg: NIU ITMO Publ., 2011, 141 p. (rus)

6. **Landsberg G.S.** *Optika [Optics]*. Moscow: Fizmatlit Publ., 2003, 848 p. (rus)

7. **Ivanova T.V., Zuyeva L.V.** Issledovaniye

sposobov diskretizatsii istochnika pri modelirovaniyefotolitograficheskogo izobrazheniya [Investigation of the sample source in the simulation of the photolithographic image]. *Opticheskiy zhurnal [Journal of Optical Technology]*, 2012, Vol. 79, No. 5, Pp. 48–52. (rus)

8. **Mozharov G.A.** *Teoriya aberratsiy opticheskikh system [The theory of the aberrations of optical systems]*. St. Petersburg: Lan Publ., 2013, 288 p. (rus)

9. **Nikulina Ye.A., Zverev V.A.** Issledovaniye vliyaniya dvulucheprelomleniya na kachestvo izobrazheniya fotolitograficheskikh sistem pri chastichno kogerentnom osveshchenii [Study of birefringence influence on image quality of photolithography systems in view of partially-coherent light source]. *Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki [Scientific and Technical Journal of Information Technologies, Mechanics and Optics]*, 2015, Vol. 15, No. 2, Pp. 196–201. (rus)

10. *Vvedeniye v tekhnologiyu CUDA – Kompyuternaya grafika i multimedia [Introduction to CUDA – Computer graphics and multimedia]*. Available: <http://cgm.computergraphics.ru/issues/issue16/cuda> (Accessed: 28.05.2015).

11. **Boreskov A.V., Kharlamov A.A.** *Osnovy raboty s tekhnologiyey CUDA [Basics CUDA technology]*, 2010, Pp. 20–33. (rus)

НИКУЛИНА Екатерина Алексеевна – аспирант кафедры прикладной и компьютерной оптики Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики.

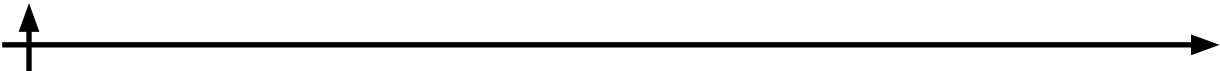
197101, Россия, Санкт-Петербург, Кронверкский пр., д. 49.

E-mail: katerina.nikulina@gmail.com

NIKULINA Ekaterina A. *St. Petersburg National Research University of Information Technologies, Mechanics and Optics.*

197101, Kronverkskiy Ave. 49, St. Petersburg, Russia.

E-mail: katerina.nikulina@gmail.com



Аппаратное обеспечение вычислительных, телекоммуникационных и управляющих систем

DOI: 10.5862/JCSTCS.229.4

УДК 621.37

А.Б. Никитин, Е.И. Хабитueva

СВЧ-МОДЕЛЬ БЕСКОРПУСНОГО РЕЗИСТОРА

A.B. Nikitin, E.I. Khabitueva

A SURFACE MOUNT CHIP RESISTOR MICROWAVE MODEL

Представлены результаты измерений в диапазоне СВЧ S -параметров бескорпусных резисторов, включенных в микрополосковую линию. Рассмотрена обобщенная модель бескорпусного резистора, предназначенная для использования при проектировании СВЧ-устройств, выполненных по гибридной технологии. На основе экспериментальных данных уточнены параметры модели, адекватно отражающей его свойства в сантиметровом диапазоне волн.

БЕСКОРПУСНОЙ РЕЗИСТОР; ЭКВИВАЛЕНТНАЯ СХЕМА; S -ПАРАМЕТРЫ; СВЧ.

An accurate model for a chip resistor is required for designing microwave devices using CAD simulation. In this connection, the measured S -parameters of chip-resistors are presented in the article. The tested device has been mounted in the gap of the microstrip line. The article discusses a generalized model of the chip resistor for use in designing various microwave devices made by hybrid technology. The proposed model has been verified based on the measurement data and adequately reflects the characteristics of the chip resistor at microwave frequencies. These results can be used for CAD simulation of microwave devices made by hybrid technology.

CHIP RESISTOR; MODEL; S -PARAMETERS; MICROWAVES.

В радиотехнических устройствах, выполненных по гибридной технологии, в качестве пассивных компонентов электронных схем широко используются так называемые SMD-компоненты (Surface Mounted Devices) – элементы для поверхностного монтажа. Выполненные в виде микрочипов, они выполняют функции конденсаторов, катушек индуктивностей, резисторов. Однако, несмотря на миниатюрность этих элементов, минимальные размеры которых для высокочастотных систем могут составлять доли миллиметра, при разработке устройств СВЧ-диапазона необходимо учитывать «паразитные» параметры таких SMD-компонентов. Их влияние в дециметровом, а тем более в сантиметровом диапазонах волн оказывается столь существенным, что

пренебрежение ими в процессе проектирования может приводить к значительным погрешностям [1–6]. Поэтому в процессе разработки гибридных СВЧ-устройств, построенных на основе SMD-компонентов, необходимы модели, адекватно отражающие импедансные характеристики в требуемом частотном диапазоне.

В отличие от конденсаторов и индуктивностей, для которых полные наборы измеренных S -параметров, как правило, доступны на сайтах ведущих производителей электронных компонентов, имеющиеся в свободном доступе данные об SMD-резисторах весьма ограничены. При этом S -параметры резисторов обычно отсутствуют, а предоставляемая в некоторых случаях информация о частотных характеристиках

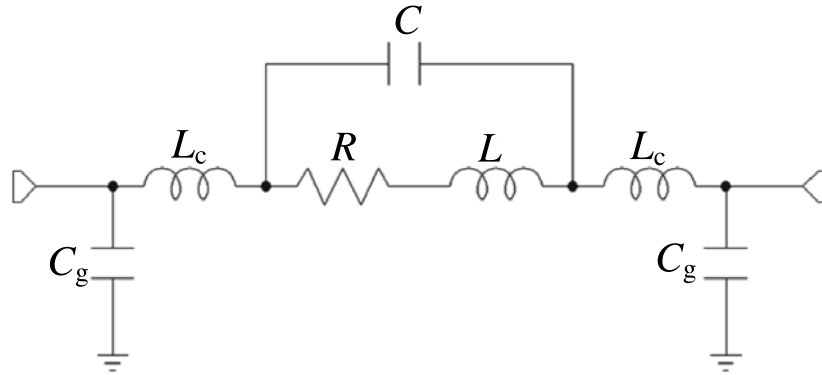


Рис. 1. Типичная эквивалентная схема SMD-резистора

компонентов носит частный характер.

Так, например, на сайте фирмы Yageo приводятся частотные зависимости модуля и фазы полного сопротивления нескольких номиналов SMD-резисторов только для типоразмеров 0805, 0603 [3]. В то же время характеристики наиболее часто применяемых в коротковолновой части дециметрового и в сантиметровом диапазонах компонентов размером 0201 отсутствуют, а для типоразмера 0402 представлены лишь зависимости модуля сопротивления.

Между тем при построении, например, твердотельных СВЧ-автогенераторов в гибридном исполнении эти SMD-компоненты широко используются в цепях автосмещения транзистора или в выходных широкополосных согласующих цепях. Поэтому при моделировании таких устройств необходимы частотные характеристики применяемых резистивных компонентов. В научно-технической литературе представлены типичные эквивалентные схемы SMD-резисторов [1, 4]. Одна из таких схем приведена на рис. 1, где введены следующие обозначения: R – номинальное сопротивление компонента; C – шунтирующая емкость; L – внутренняя индуктивность; L_c – внешняя индуктивность контактов; C_g – внешняя емкость на землю [4].

Практическое применение таких схем при разработке СВЧ-устройств требует наличия достоверной информации о значениях всех параметров модели. При этом использование в процессе моделирования эквивалентной схемы компонента (рис. 1), в которой значения параметров соответствуют

публикуемым в научно-технической литературе данным [4], приводит, как показали результаты экспериментов, к существенным расхождениям с характеристиками SMD-резисторов, измеренными в сантиметровом диапазоне волн. Наблюдаемые расхождения могут быть связаны как с неточными данными о значениях параметров модели (R , C , L , L_c , C_g), так и с возможными особенностями процедуры измерений, на основе которых строилась представленная эквивалентная схема (рис. 1).

Поэтому при создании ряда СВЧ-устройств в гибридном исполнении требуются достоверные экспериментальные данные об импедансных характеристиках используемых SMD-резисторов или модель, адекватно отражающая их частотные свойства и позволяющая осуществлять процесс разработки с помощью современных программ автоматизированного проектирования. В данной статье рассматриваются результаты измерения характеристик

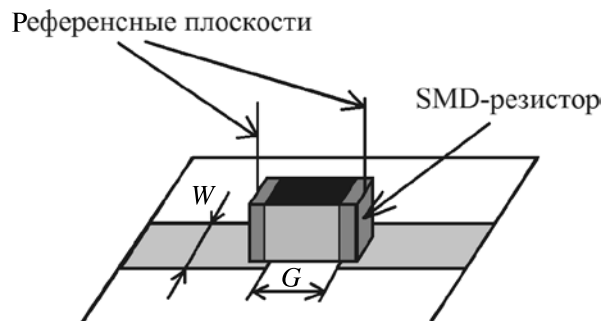


Рис. 2. Схема включения измеряемого компонента в разрыв микрополосковой линии

SMD-резисторов в сантиметровом диапазоне волн, на основе которых разработана универсальная модель резистивного компонента, включаемого в разрыв микрополосковой линии.

В ходе экспериментальных исследований были измерены в диапазоне частот от 3 ГГц до 15 ГГц S -параметры SMD-резисторов компании Yageo номиналом от единиц Ом до нескольких кОм. Экспериментальное исследование характеристик электронных компонентов, включенных в разрыв микро-

полосковой линии (МПЛ), производилось с помощью векторного анализатора цепей E5071C ENA Series Network Analyzer компании Agilent (Keysight) Technologies [7]. На рис. 2 схематично представлен внешний вид тестовых плат, изготовленных из фольгированного СВЧ-материала Rogers RO4350В толщиной 0,254 мм, с включенным в МПЛ SMD-резистором.

В процессе измерений использовались следующие параметры МПЛ: $W = 0,54$ мм и $G = 0,3$ мм.

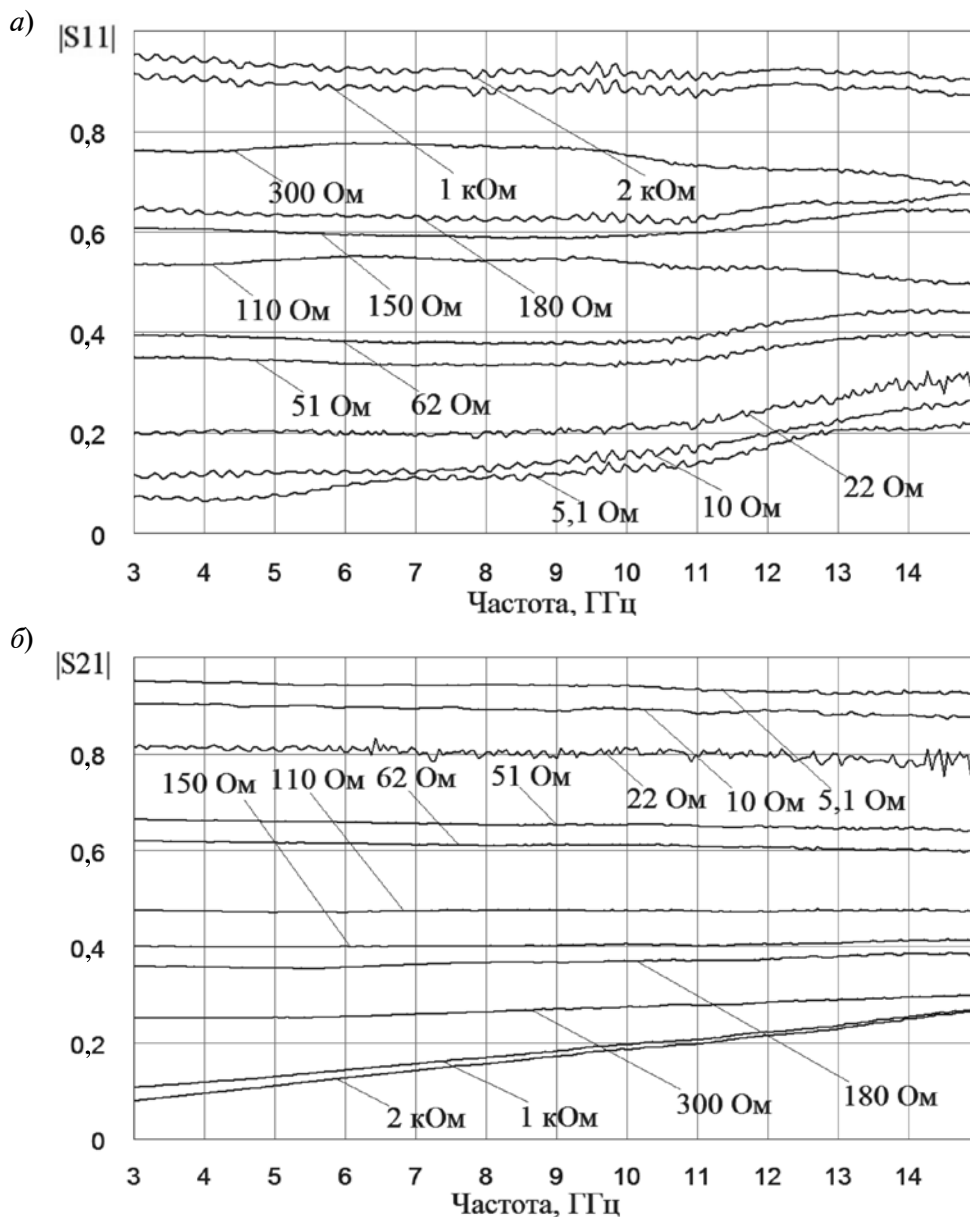


Рис. 3. Измеренные зависимости модулей S -параметров SMD-резисторов

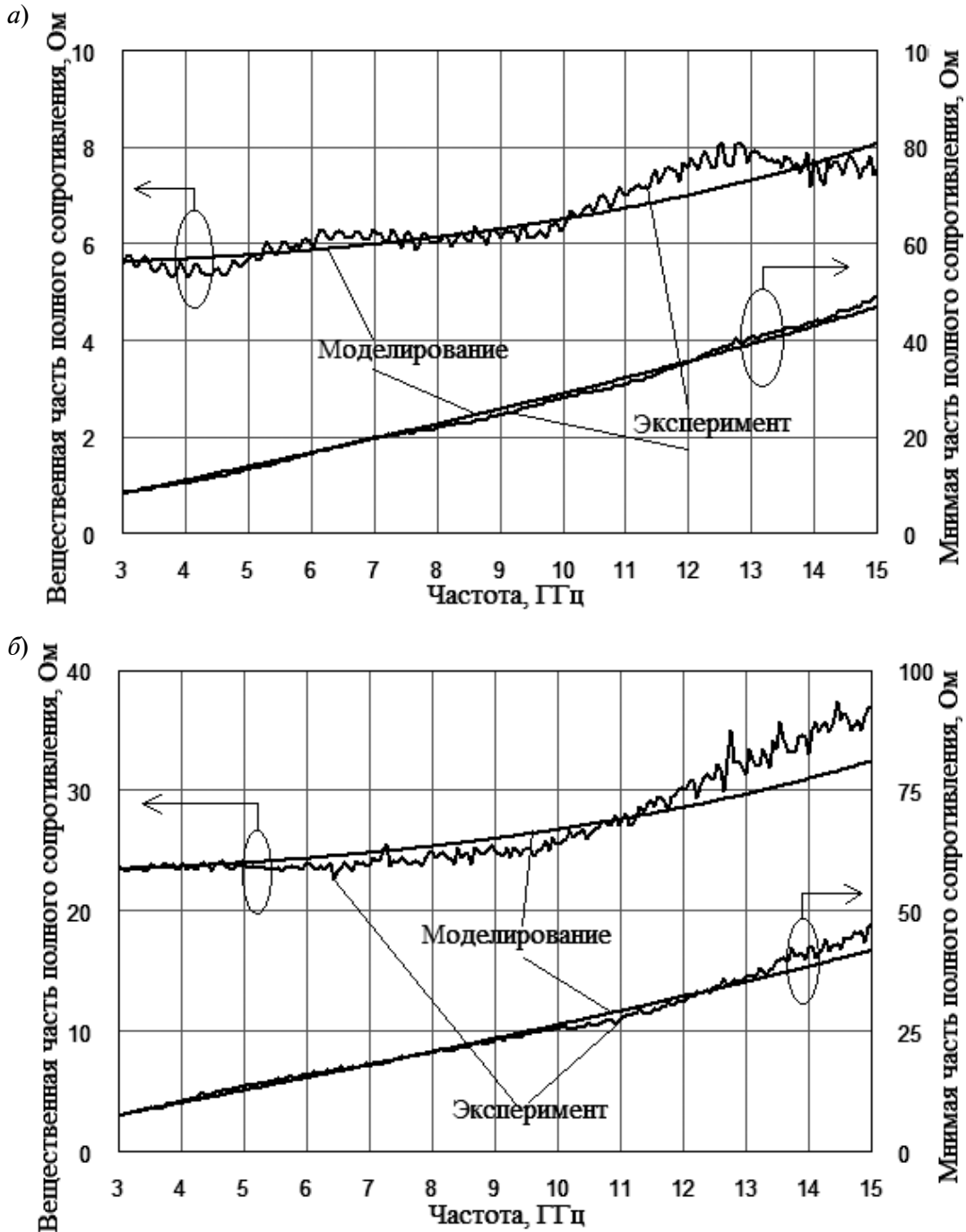


Рис. 4. Импедансные характеристики SMD-резисторов номиналом 5,1 Ом (а) и 22 Ом (б)

При обработке результатов измерений для приведения экспериментальных данных к референсным плоскостям и исключения влияния неоднородностей подводящих линий и переходов использовались S-параметры эталонных плат с цельной МПЛ, характеристики которой

были идентичны характеристикам линий в тестовых платах с установленными SMD-компонентами. Плоскости приведения результатов измерения (референсные плоскости) выбраны на торцах исследуемых компонентов (рис. 2).

В качестве примера на рис. 3 пред-

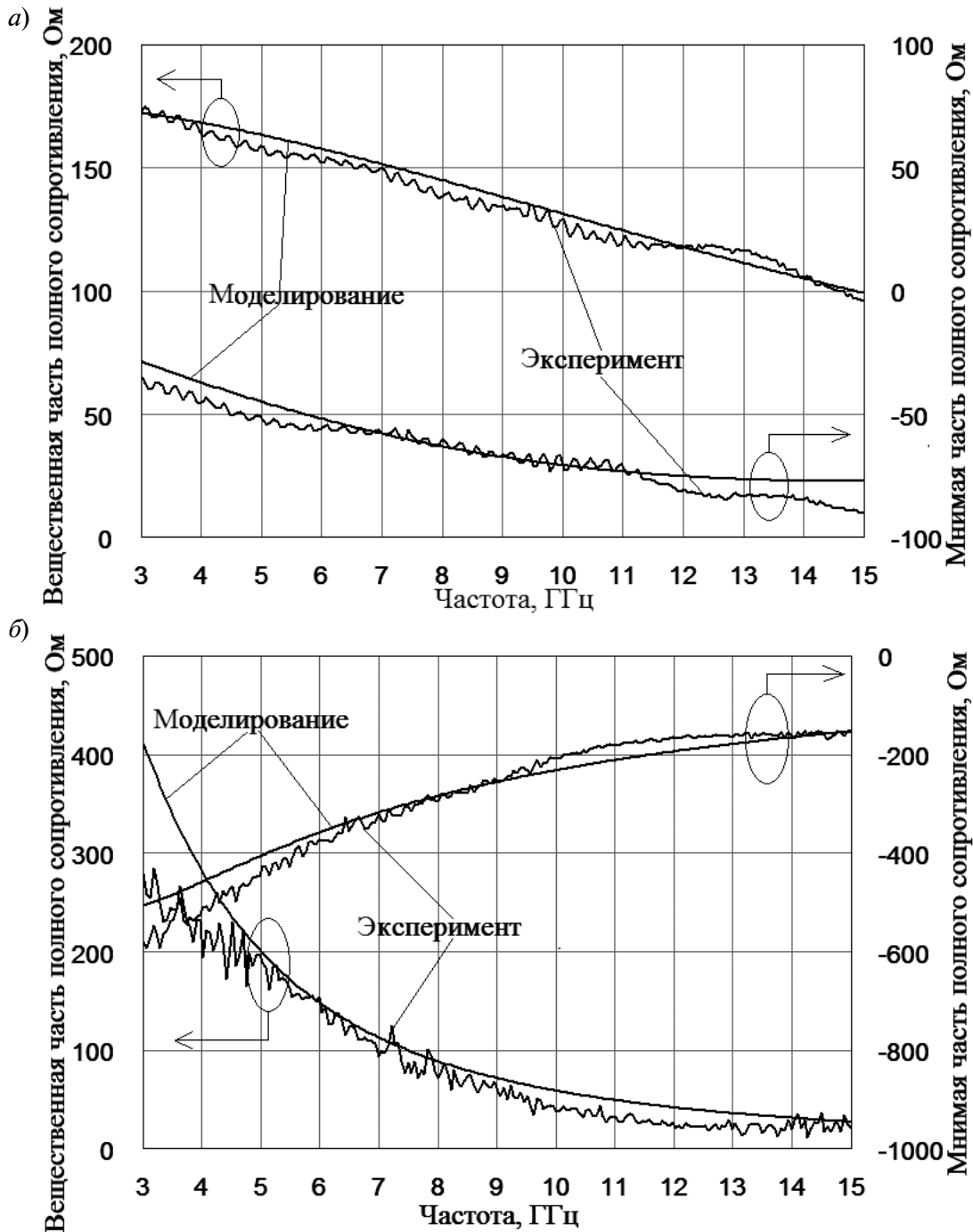


Рис. 5. Импедансные характеристики SMD-резисторов номиналом 180 Ом (а) и 1,3 кОм (б)

ставлены частотные зависимости модулей S -параметров серии резисторов типоразмера 0201 номиналом от 5,1 Ом до 2 кОм ($|S_{11}|$ на рис. 3 а и $|S_{21}|$ на рис. 3 б).

Полученные в результате измерений частотные зависимости S -параметров были использованы для нахождения импе-

дансных характеристик исследуемых компонентов. На рис. 4 и 5 приведены зависимости от частоты вещественной и мнимой частей полного сопротивления SMD-резисторов типоразмера 0201 нескольких номиналов.

Как следует из представленных зависи-

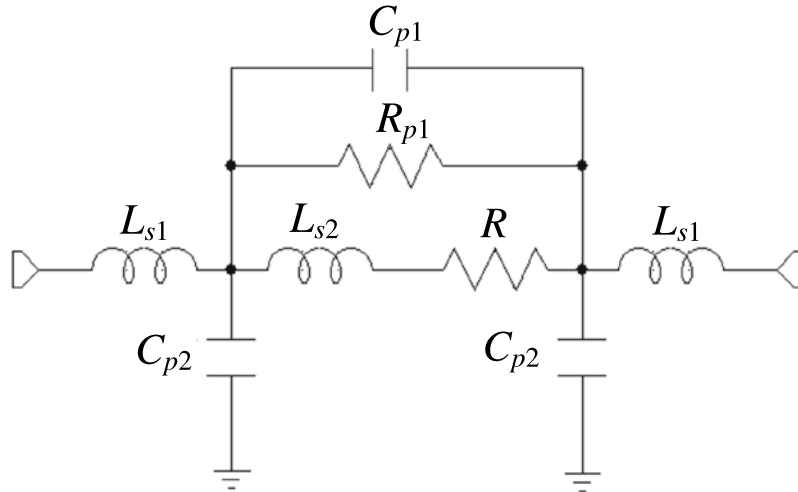


Рис. 6. Эквивалентная схема SMD-резистора

мостей, влияние дополнительных параметров исследуемых компонентов в рассматриваемом диапазоне оказывается весьма существенным. Например, на частоте 10 ГГц для резистора номиналом 22 Ом вещественная составляющая полного сопротивления оказывается равной мнимой (рис. 4 б), а для резистора номиналом 1,3 кОм мнимая составляющая по абсолютной величине в несколько раз превосходит вещественную (рис. 5 б).

Полученные экспериментальные данные были использованы для построения универсальной модели, адекватно отражающей частотные характеристики SMD-резисторов различных номиналов, включенных в разрыв МПЛ. Соответствующая эквивалентная схема SMD-резистора приведена на рис. 6.

В данной модели символом R обозначен номинал исследуемого резистора, L_{s1} и L_{s2} представляют собой соответственно индуктивность выводов и индуктивность самого резистивного слоя, под R_{p1} подразумевается сопротивление защитного покрытия, а символы C_{p1} и C_{p2} учитывают соответственно емкость между выводами корпуса и ем-

кость между выводами и землей.

Значения параметров (L_{s1} , L_{s2} , R_{p1} , C_{p1} , C_{p2} , рис. 6) определялись методами параметрической оптимизации на основе сравнения в исследуемом диапазоне частот S -параметров модели с результатами измерений тестовых образцов SMD-резисторов. Найденные значения представлены в таблице для набора компонентов типоразмера 0201 с номиналами, принадлежащими диапазону от 5 Ом до 2 кОм.

Импедансные характеристики нескольких SMD-резисторов соответствующих номиналов, рассчитанные с использованием данной эквивалентной схемы (рис. 6), приведены на графиках (см. рис. 4, 5).

Отличие измеренных значений модуля полного сопротивления резисторов от величин, полученных для их моделей, в различных точках частотного диапазона меняется от нескольких единиц до двух десятков процентов. При этом, например, на частоте 9 ГГц максимальная разница между данными эксперимента и моделирования наблюдается для резистора 2 кОм и не превышает 27 %. Для подавляющего числа номиналов

Параметры модели

L_{s1} , нГн	L_{s2} , нГн	C_{p1} , пФ	C_{p2} , пФ	R_{p1} , кОм
0,098	0,239	0,026	0,037	5,725

из диапазона 5÷180 Ом максимальное отличие не превосходит 7 %.

Таким образом, в результате проделанной работы получены экспериментально снятые в сантиметровом диапазоне импедансные характеристики SMD-резисторов,

включаемых в разрыв микрополосковой линии. Уточнена эквивалентная схема SMD-резисторов номиналом от 5,1 Ом до 2 кОм, адекватно описывающая данные измерений. Полученные результаты могут использоваться при разработке СВЧ-устройств в гибридном исполнении.

СПИСОК ЛИТЕРАТУРЫ

1. **Bahl I.** *Lumped Elements for RF and Microwave Circuits*. Artech House, Inc., 2003. 488 p.
2. **Stauffer G.H.** Finding the Lumped Element Varactor Diode Model // *High Frequency Electronics*. 2003. Vol. 2. No. 6. Pp. 22–28.
3. Thick film chip resistors. Datasheet [электронный ресурс] / URL: <http://www.yageo.com> (дата обращения: 11.09.2015).
4. Frequency Response of Thin Film Chip Resistors. Technical note [электронный ресурс] / URL: <http://www.vishay.com/docs/60107/freqresp.pdf> (дата обращения: 11.09.2015).
5. **Дидилев С.** Особенности применения и

критерии выбора конденсаторов и резисторов для работы в цепях ВЧ/СВЧ // *Компоненты и технологии*. 2005. № 5. С. 38–44.

6. **Мальшев В.М., Матвеев Ю.А., Никитин А.Б., Худяков А.В.** Модель варикапа для разработки сверхширокополосных перестраиваемых генераторов СВЧ // *Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление*. СПб.: Изд-во Политехн. ун-та, 2014. № 2(193). С. 55–60.

7. Keysight Technologies. E5071C ENA Network Analyzer. Datasheet [электронный ресурс] / URL: <http://www.keysight.com> (дата обращения: 11.09.2015).

REFERENCES

1. **Bahl I.** *Lumped Elements for RF and Microwave Circuits*. Artech House, Inc., 2003, 488 p.
2. **Stauffer G.H.** Finding the Lumped Element Varactor Diode Model. *High Frequency Electronics*, 2003, Vol. 2, No. 6, Pp. 22–28.
3. *Thick film chip resistors. Datasheet*. Available: <http://www.yageo.com> (Accessed: 11.09.2015).
4. *Frequency Response of Thin Film Chip Resistors. Technical note*. Available: <http://www.vishay.com/docs/60107/freqresp.pdf> (Accessed: 11.09.2015).
5. **Didilev S.** Osobennosti primeneniya i kriterii vybora kondensatorov i rezistorov dlya raboty v tsepyah VCh/SVCh [Features of the application and selection criteria capacitors and resistors for use in the circuits of RF / Microwave]. *Komponenty*

i tehnologii [Components & Technologies], 2005, No. 5, Pp. 38–44. (rus)

6. **Malyshev V.M., Matveev Yu.A., Nikitin A.B., Khudyakov A.V.** Model varikapu dlya razrabotki sverhshirokopolosnykh perestraivaemykh generatorov SVCh [Varactor diode model used to design wideband microwave voltage-controlled oscillators]. *Nauchno-tehnicheskie vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravlenie [St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control System]*. St. Petersburg: SPbGPU Publ., 2014, No. 2(193), Pp. 55–60. (rus)

7. *Keysight Technologies. E5071C ENA Network Analyzer. Datasheet*. Available: <http://www.keysight.com> (Accessed: 11.09.2015).

НИКИТИН Александр Борисович — доцент кафедры радиотехники и телекоммуникаций Института физики, нанотехнологий и телекоммуникаций Санкт-Петербургского политехнического университета Петра Великого, кандидат технических наук.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: nikitin@mail.spbstu.ru

NIKITIN Aleksandr B. Peter the Great St. Petersburg Polytechnic University.

195251, Politehnicheskaya Str. 29, St. Petersburg, Russia.

E-mail: nikitin@mail.spbstu.ru

ХАБИТУЕВА Екатерина Исаковна – студентка магистратуры Санкт-Петербургского политехнического университета Петра Великого.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: basilliounderground@mail.ru

KNABITUEVA Ekaterina I. *Peter the Great St. Petersburg Polytechnic University.*

195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.

E-mail: basilliounderground@mail.ru

**РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ
РАЗМЕЩЕНИЯ ОБЪЕКТОВ ТЕХНОГЕННОЙ ОПАСНОСТИ
С ИСПОЛЬЗОВАНИЕМ НЕЧЕТКОЙ ЛОГИКИ**

O.I. Khristodulo, I.F. Salimzynov, N.R. Gareeva

**THE DEVELOPMENT OF AN INFORMATION SYSTEM FOR POSITIONING
TECHNOGENIC HAZARD OBJECTS USING FUZZY LOGIC**

Описано применение нечеткой логики в информационных системах. Изучены предпосылки использования нечеткой логики, показана математическая модель обработки данных, отображен пример работы информационной системы и сравнительный анализ эффективности ее использования.

СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ; ИСПОЛЬЗОВАНИЕ НЕЧЕТКОЙ ЛОГИКИ В ГИС; УПРАВЛЕНИЕ ТЕХНОГЕННОЙ БЕЗОПАСНОСТЬЮ; МНОГОКРИТЕРИАЛЬНАЯ МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДЛЯ ОБРАБОТКИ ДАННЫХ.

This article describes the development of a technogenic hazard management system aimed at landfill positioning. This information system uses fuzzy logic in its operations. The article shows that providing technogenic safety is a fundamental problem in all industrialized countries. One of the main tasks in technogenic safety is landfill positioning, and one of the convenient ways of solving it is using fuzzy logic calculations. The article describes the prerequisites for using fuzzy logic, shows a multicriteria mathematical model for data processing which is normally used in landfill positioning. This model contains a set of linguistic variables describing the input conditions of the analysis and a fuzzy rules base combining these variables. Another model used in the information system is a data processing model developed in ArcGIS Model Builder. The article also presents a logic structure of the system, an example of an information system working and a comparative analysis of the efficiency of its use.

INFORMATION PROCESSING SYSTEMS; USING FUZZY LOGIC IN GIS; TECHNOGENIC SAFETY MANAGEMENT; MULTICRITERIA MATHEMATICAL MODEL FOR DATA PROCESSING.

Одна из фундаментальных проблем промышленно развитых стран – обеспечение приемлемого уровня техногенной безопасности. Ошибки, допускаемые на всех уровнях управления техногенной безопасностью, создают угрозу здоровью и жизни населения, являются источником социальной напряженности.

Управление техногенной опасностью в

настоящий момент является одним из приоритетных направлений в работе международного сообщества. Растущие объемы промышленного производства провоцируют симметричный рост экологических и техногенных рисков и угроз. Стремление контролировать этот рост привело к наблюдаемому в последние десятилетия экспоненциальному развитию различных про-

грамм и научных разработок, связанных с техногенной безопасностью.

Объекты техносферы, несущие в себе потенциальную угрозу, разнятся по своему составу, уровню опасности и масштабам распространенности. Помимо действующих промышленных производств, являющихся источниками активной техногенной опасности, необходимо выделить захоронения отходов, которые несут пассивные, нередко отложенные во времени риски. Долговременное воздействие таких захоронений на окружающую среду может спровоцировать серьезные экологические последствия. Будучи промышленно развитым государством, Российская Федерация не является исключением и аккумулирует на своей территории значительное количество промышленных и бытовых отходов (согласно [1], накопленное количество отходов на территории страны превышает 94 млрд т).

В данных условиях важнейшим фактором, влияющим на техногенную опасность, становится правильное размещение объектов захоронения отходов. Однако задача размещения отличается определенной спецификой в связи с неточностью и недостаточностью исходной информации, что отражается на сложности в формулировании достоверных выводов. Разработка методов, позволяющих получать адекватные результаты моделирования о возможном размещении объектов отходов с учетом неопределенности исходной информации и требований к их местоположению, имеет важное практическое значение. Наиболее перспективным при работе в данном направлении видится привлечение экспертных знаний в рамках нечеткой логики.

Использование нечеткой логики при размещении объектов техногенной опасности

При определении территорий, пригодных для размещения объектов техногенной опасности, учитывается влияние ряда факторов, в соответствии с которыми выносятся решение о пригодности территории к применению по целевому назначению. Однако использование строгих правил при классификации территорий снижает гиб-

кость и оказывает ограничивающий эффект на процесс принятия управленческих решений, что может привести к росту как социально-экономических, так и техногенных рисков. Более целесообразной является классификация на основе теории нечетких множеств.

Основоположниками теории нечетких множеств являются Л. Заде и Д. Клауа. Авторы расширили понятие множества, допустив, что функция принадлежности элемента множеству может принимать любые значения в интервале от нуля до единицы, а не только нуль или единица [2, 3, 8]. Использование понятия нечеткого множества в задачах классификации обеспечивает большую гибкость при отнесении территории к тому или иному классу, если границы между этими классами имеют переходный характер.

Ядром теории служит понятие нечеткого множества a с функцией принадлежности $\mu_a(r)$ и областью значений $[0;1]$. В данном случае $r \in U$ – множество входных четких значений. Чем больше $\mu_a(r)$, тем в большей степени элемент универсального множества соответствует свойствам нечеткого множества. Если U – непрерывное множество, то нечеткое множество a записывается следующим образом:

$$a = \int_U (\mu_a(r) / r) dr. \quad (1)$$

Размещение объектов техногенной опасности осуществляется путем математической формализации нечетких данных, основанных на экспертных методах оценки. Оптимальная позиция объекта определяется характеристикой, называемой *лингвистической переменной* [4]. Сформулируем постановку задачи для определения степени пригодности территории к размещению на ней объектов техногенной опасности.

Пусть дан набор критериев $k_i = [k_1, \dots, k_n]$, определяющих оптимальный участок для размещения объектов техногенной опасности. Каждый критерий описывается лингвистической переменной, представленной в виде:

$$K_i = \langle A_i, U_i, T_i \rangle, \quad (2)$$

где A_i – наименование лингвистической переменной; U_i – универсальное множество входных значений; T_i – набор возможных значений лингвистической переменной, которые в рамках исследования размещения объектов техногенной опасности представлены как «Абсолютно пригоден», «Пригодно» и др.

Требуется найти выходной параметр, определяющий степень пригодности территории к размещению объекта техногенной опасности, который описывается лингвистической переменной y . Необходимо определить влияние каждой из пяти входных лингвистических переменных на выбор участка территории, для чего следует сформировать базу правил нечеткого вывода [9]. Такие правила устанавливаются при помощи зависимостей, представляющих собой логические высказывания вида ЕСЛИ A , то B :

$$P : A \rightarrow B, \quad (3)$$

где P – условие применимости ядра (обычно представляет собой логическое выражение).

При проведении пространственного анализа на основе нечеткой логики важным аспектом является выбор функции принадлежности и ее параметров для нечетких множеств [5, 10]. В практической деятельности применяются следующие виды функций принадлежности:

треугольные и трапециевидные (линейно-кусочные);

нелинейные (функция Гаусса, сигмоидальная функция, сплайн);

LR-представление функций принадлежности.

Использование этих функций сокращает численные расчеты и, соответственно, вычислительные ресурсы, необходимые для хранения отдельных значений функции принадлежности.

Многокритериальная математическая модель размещения объектов техногенной опасности на основе нечеткой логики

Согласно нормативным актам, в которых указаны основные требования к месту расположения полигонов отходов, на вы-

бор территории для размещения полигона накладывается ряд ограничений [6]:

расстояние до населенных пунктов и открытых водоемов устанавливается с учетом конкретных местных условий, но не менее 3000 м;

расстояние от границ леса и лесопосадок, не предназначенных для использования в рекреационных целях, – не менее 50 м;

глубина залегания грунтовых вод – не менее 2 м от нижнего уровня захороненных отходов;

размещение полигонов не допускается в зонах активного карста.

Отметим, что перечисленные требования являются минимально необходимыми. На выбор территории также влияют долгосрочные перспективы эксплуатации объекта захоронения отходов, такие как изменение состава хранимых отходов, увеличение площади полигона и др.

Для достижения максимальных эксплуатационных возможностей математической модели, а также с целью обеспечения корректности нечеткой базы знаний, к разработке модели были привлечены специалисты отдела экологической безопасности Министерства природопользования и экологии Республики Башкортостан. Группа экспертов, состоящая из сотрудников отдела, работала в соответствии с прямым методом построения [11], согласно которому эксперт непосредственно задает значение функции принадлежности (степень соответствия элемента r) для каждого значения исследуемого параметра.

Значения функции принадлежности вычислялись по формуле:

$$\mu_A = \frac{n_A}{n}, \quad (4)$$

где n – число экспертов, участвовавших в эксперименте; n_A – число экспертов, проголосовавших за соответствие элемента r нечеткому множеству A .

По полученным в результате экспертного анализа значениям функции фиксируются характерные точки. Следующий этап – аппроксимация подходящей типовой функцией по характерным точкам. Нечеткие подмножества, полученные в результате

этих операций, ассоциируются с определенным термом из формализованного экспертами терм-множества.

В разработанной модели использовались следующие типы функций: S-образные, Z-образные и сигмоидальные функции

Требования к месту расположения полигонов отходов были выражены через следующие лингвистические переменные:

- расстояние до грунтовых вод;
- расстояние до карстовых пород;
- расстояние до населенных пунктов;
- расстояние до водных объектов;
- расстояние до лесного массива.

Фрагмент модели, соответствующей лингвистической переменной «Глубина залегания грунтовых вод», описан выражениями (5), (6). Лингвистическая переменная обладает тремя значениями, или термами – «Недопустимо близко», «Близко» и «Далеко», – каждый из которых описывается соответствующей функцией принадлежности. Таким образом, из рис. 1 видно, что степень принадлежности, например, терму «Близко» снижается при изменении глубины залегания грунтовых вод с 2 до 4 м. Соответственно, в этом же диапазоне возрастает степень принадлежности переменной к терму «Далеко».

Переменная «Расстояние до грунто-

вых вод»:

$$A_1^{гв} \in T_1^{гв} : T_1^{гв} = \left\{ \begin{array}{l} a_{1,1} - \text{Недопустимо близко,} \\ a_{1,2} - \text{Близко,} \\ a_{1,3} - \text{Далеко} \end{array} \right\}, \quad (5)$$

$$\mu_{1,1} = \begin{cases} 1, & r_1 \leq 2, \\ 0, & r_1 > 2, \end{cases}$$

$$\mu_{1,2} = \begin{cases} \frac{1}{2} + \frac{1}{2} \cos\left(\frac{r_1 - 2}{2} \pi\right), & 2 < r_1 \leq 4, \\ 0, & r_1 > 4 \end{cases} \quad (6)$$

$$\mu_{1,3} = \begin{cases} 0, & r_1 < 2, \\ \frac{1}{2} + \frac{1}{2} \cos\left(\frac{r_1 - 4}{2} \pi\right), & 2 \leq r_1 \leq 4, \\ 1, & r_1 > 4 \end{cases}$$

Другой критерий, описанный в рамках многокритериальной математической модели, «Расстояние до населенных пунктов» формализован формулами (7), (8). В отличие от предыдущего критерия, отображенного на рис. 1, данный критерий обладает четырьмя лингвистическими термами (рис. 2).

Переменная «Расстояние до населенных пунктов»:

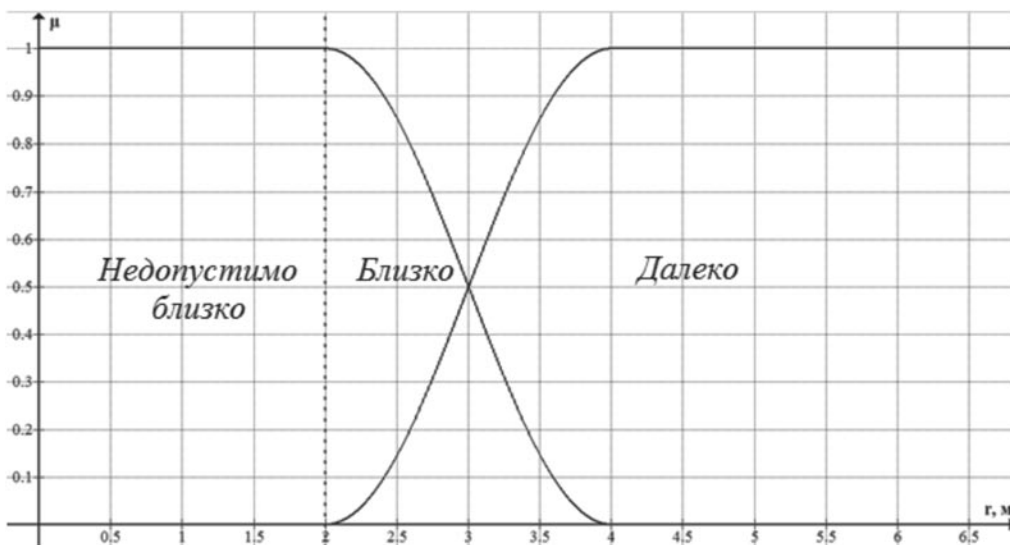


Рис. 1. Функция принадлежности для термов лингвистической переменной «Расстояние до грунтовых вод»

$$A_3^{/nn} \in T_3^{/nn} : T_3^{/nn} = \left\{ \begin{array}{l} a_{3,1} - \text{Недопустимо близко,} \\ a_{3,2} - \text{Близко,} \\ a_{3,3} - \text{Недалеко,} \\ a_{3,4} - \text{Далеко} \end{array} \right\}. \quad (7)$$

$$\mu_{3,1} = \begin{cases} 1, & r_3 \leq 3000, \\ 0, & r_3 > 3000 \end{cases}$$

$$\mu_{3,2} = \frac{1}{1 + e^{(0,012(r_3 - 3500))}}, \quad (8)$$

$$\mu_{3,3} = \begin{cases} \frac{1}{1 + e^{(-0,012(r_3 - 3500))}}, & r_3 \leq 4000, \\ \frac{1}{1 + e^{(0,012(r_3 - 4500))}}, & r_3 > 4000 \end{cases}$$

$$\mu_{3,4} = \frac{1}{1 + e^{(-0,012(r_3 - 4500))}}.$$

Полная многокритериальная математическая модель размещения объектов техногенной опасности на основе нечеткой логики приведена в [7]. Необходимо отметить, что разработанная модель позволяет вводить неограниченное число критериев, например, преобладающее направление ветра, уклон поверхности и др.

На основе экспертных знаний была сформирована база правил, определяющая

степень пригодности анализируемой территории для размещения на ней объекта техногенной опасности для захоронения отходов (см. табл.).

Для автоматизации обработки данных разработана модель в среде ArcGIS Model Builder. Алгоритм обработки основан на процедуре нечеткого логического вывода и состоит из следующих этапов:

- преобразование векторных поверхностей в растры;
- объединение части растров (необходимо для снижения временных затрат на обработку данных);
- фаззификация (перевод характеристик в нечеткую форму с помощью функций принадлежности);
- агрегирование (определение степени истинности условия по каждому правилу);
- активизация подусловий (объединение полученных на этапе агрегирования степеней истинности);
- дефаззификация (преобразование нечетких данных обратно в четкую форму);
- представление результатов.

На рис. 3 представлена общая структура модели и отдельно вынесен этап представления результатов. На этом этапе формируются поверхности, соответствующие термам выходной лингвистической пере-

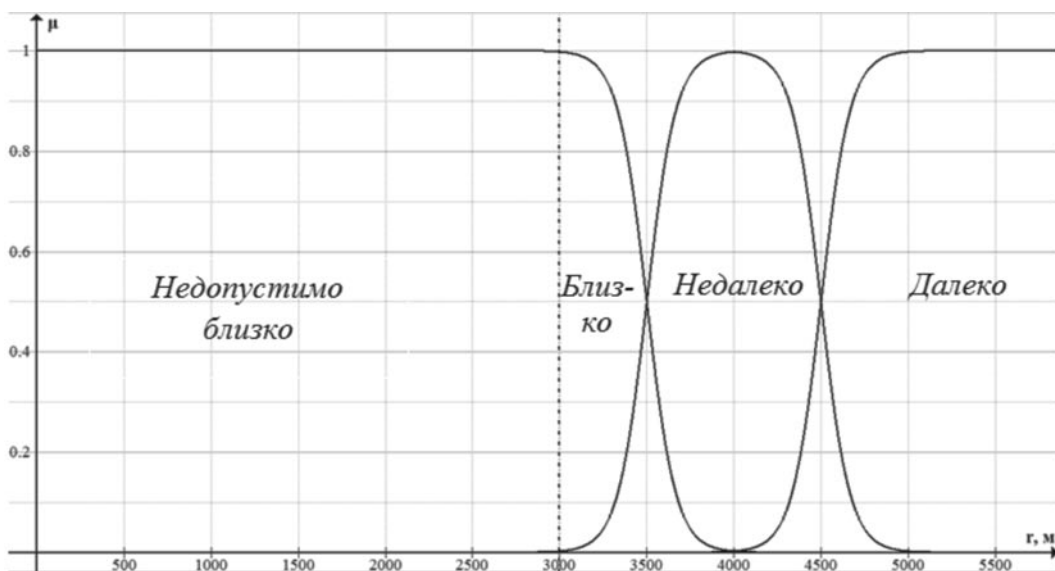


Рис. 2. Функция принадлежности для термов лингвистической переменной «Расстояние до населенных пунктов»

База правил

	ЕСЛИ $r_1 = A_1$	И $r_2 = A_2$	И $r_3 = A_3$	И $r_4 = A_4$	И $r_5 = A_5$	ТО $y = B_i$
1	далеко	далеко	далеко	далеко	далеко	Абсолютно пригодна
2	далеко	далеко	далеко	далеко	близко	
3	далеко	далеко	далеко	близко	далеко	
4	далеко	далеко	близко	далеко	далеко	
5	далеко	далеко	недалеко	далеко	далеко	
6	далеко	далеко	недалеко	далеко	близко	Пригодна
7	далеко	далеко	недалеко	близко	далеко	
8	далеко	близко	недалеко	далеко	далеко	
9	близко	далеко	недалеко	далеко	далеко	
10	далеко	далеко	далеко	близко	близко	
11	далеко	близко	далеко	далеко	далеко	
12	далеко	близко	далеко	далеко	близко	
13	далеко	близко	далеко	близко	далеко	
14	близко	далеко	далеко	далеко	далеко	
15	близко	далеко	далеко	далеко	близко	
16	близко	далеко	далеко	близко	далеко	
17	близко	близко	далеко	далеко	далеко	
18	далеко	далеко	близко	далеко	близко	
19	далеко	далеко	близко	близко	далеко	
20	далеко	близко	близко	далеко	далеко	
21	близко	далеко	близко	далеко	далеко	Недостаточно пригодна
22	далеко	близко	далеко	близко	близко	
23	близко	далеко	далеко	близко	близко	
24	близко	близко	далеко	далеко	близко	
25	близко	близко	далеко	близко	далеко	
26	далеко	далеко	близко	близко	близко	
27	далеко	близко	близко	далеко	близко	
28	далеко	близко	близко	близко	далеко	
29	близко	далеко	близко	далеко	близко	
30	близко	далеко	близко	близко	далеко	
31	близко	близко	близко	далеко	далеко	
32	далеко	далеко	недалеко	близко	близко	
33	далеко	близко	недалеко	далеко	близко	
34	далеко	близко	недалеко	близко	далеко	
35	близко	далеко	недалеко	далеко	близко	
36	близко	далеко	недалеко	близко	далеко	
37	близко	близко	недалеко	далеко	далеко	
38	близко	близко	далеко	близко	близко	
39	далеко	близко	близко	близко	близко	
40	близко	далеко	близко	близко	близко	

	ЕСЛИ $r_1 = A$	И $r_2 = A_2$	И $r_3 = A_3$	И $r_4 = A_4$	И $r_5 = A_5$	ТО $y = B$
41	близко	близко	близко	далеко	близко	Непригодна
42	близко	близко	близко	близко	далеко	
43	близко	близко	близко	близко	близко	
44	близко	близко	недалеко	близко	близко	
45	далеко	близко	недалеко	близко	близко	
46	близко	далеко	недалеко	близко	близко	
47	близко	близко	недалеко	далеко	близко	
48	близко	близко	недалеко	близко	далеко	

менной и характеризующие степень пригодности территории для размещения объектов техногенной опасности.

Информационная система размещения объектов техногенной опасности в условиях неопределенности

В целях автоматизации процесса определения степени пригодности территории для размещения объектов техногенной опасности разработана информационная система, в работе которой применяется нечеткая логика. Данная информационная система решает следующие задачи:

определение степени пригодности выбранного участка территории для размещения объектов техногенной опасности на основании заданных критериев;

подготовка отчетной документации в виде карты и ее экспорт в пригодный для печати формат;

автоматизация процесса определения степени пригодности выбранного участка территории для размещения объектов техногенной опасности путем снижения трудоемкости осуществляемых операций;

снижение требований к квалификации специалистов, осуществляющих анализ.

Структура информационной системы включает в себя базы пространственной и атрибутивной информации, содержащие общегеографическую информацию, необходимую для анализа, данные об объектах техногенной опасности, а также базу нечетких знаний (рис. 4).

В качестве примера работы информационной системы на рис. 5 отображен участок территории, на котором был осуществлен анализ степени пригодности территории для размещения объектов твердых бытовых отходов. В результате автоматизированной обработки исходных данных получены ито-

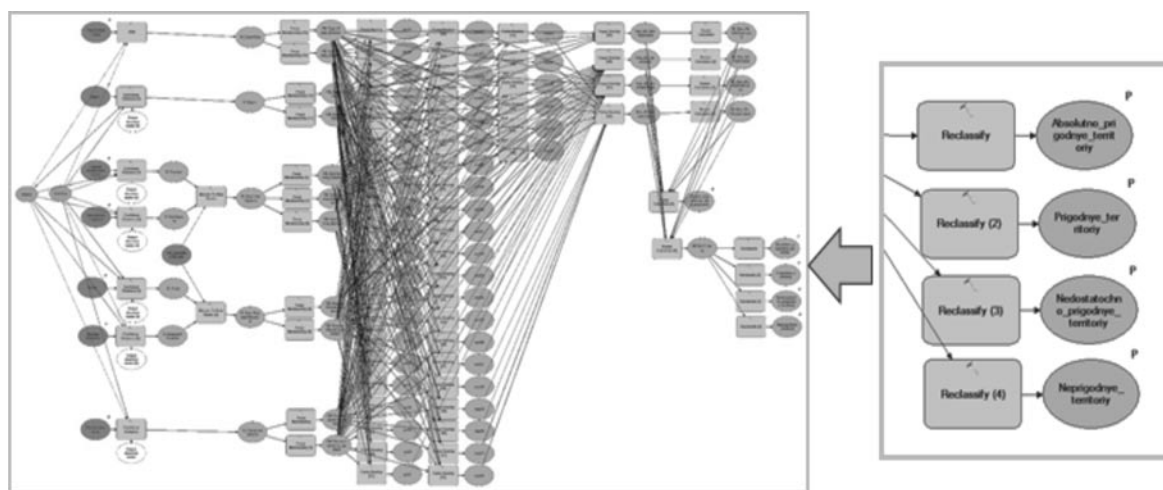


Рис. 3. Модель обработки данных в ARCGIS MODEL BUILDER

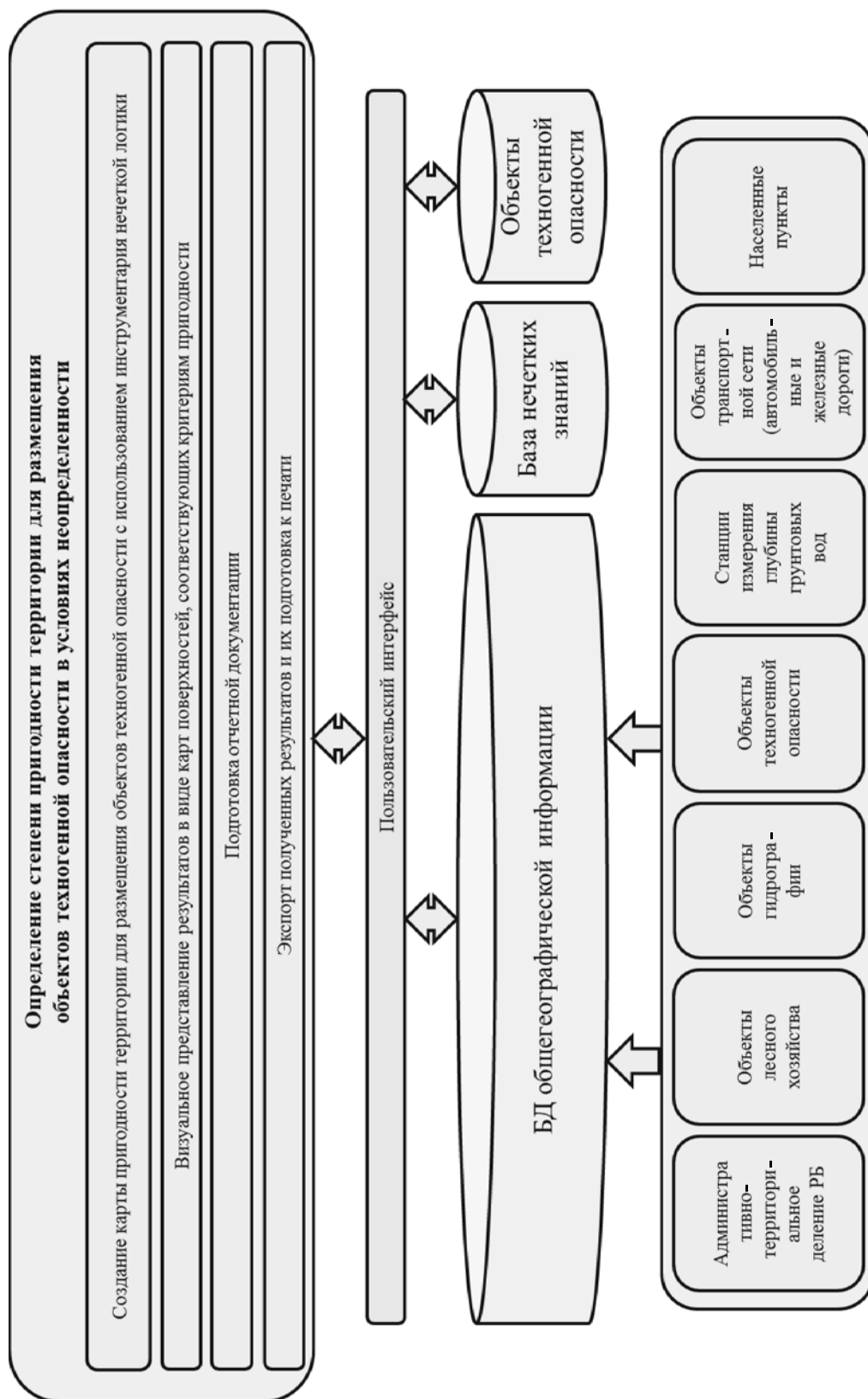


Рис. 4. Логическая структура информационной системы размещения объектов техногенной опасности

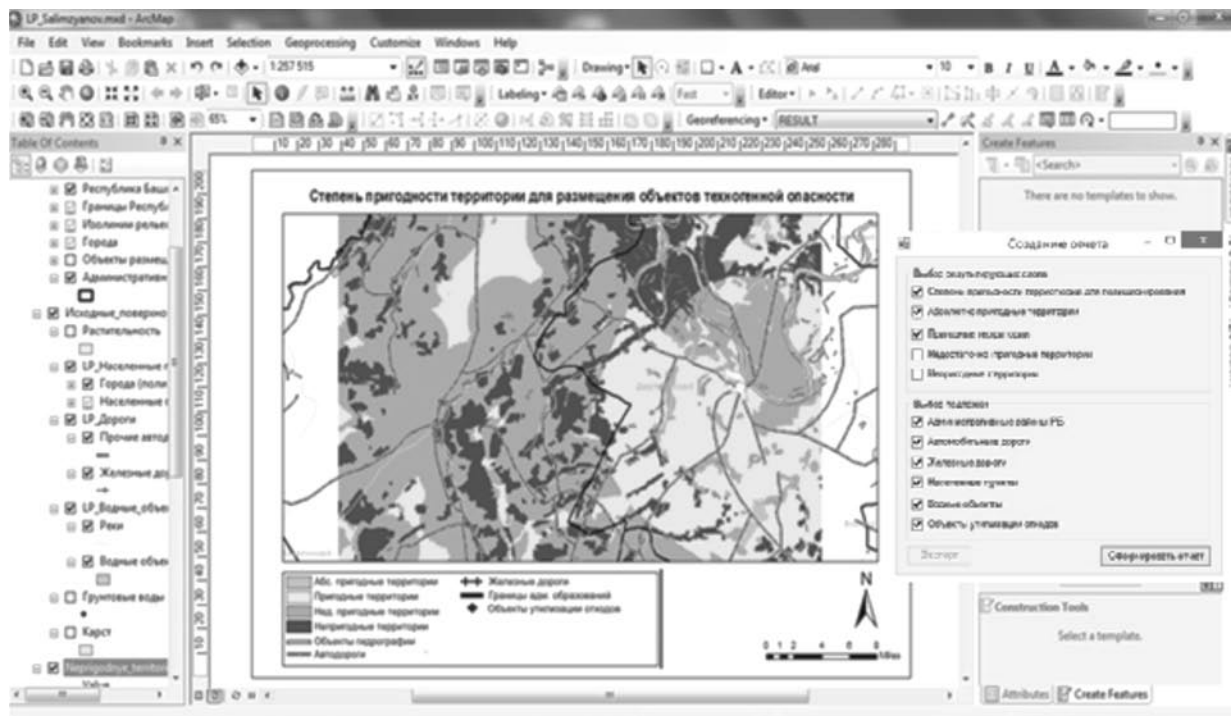


Рис. 5. Пример анализа степени пригодности территории для размещения объектов твердых бытовых отходов

говые растры, отображающие соответствие участков территории соответствующим термам (критериям) выходной лингвистической переменной, при этом формируется карта с легендой и условными обозначениями. Например, светло-зеленые участки территории являются наиболее пригодными для размещения объектов техногенной опасности. Кроме того, интерфейс системы позволяет управлять отображением отдельных результирующих поверхностей (например, если требуется отобразить на карте только непригодные для размещения объектов техногенной опасности территории) и слоев подложки, таких как дороги, водные объекты и др.

На основании данных, полученных от специалистов Минэкологии РБ, — лиц, потенциально заинтересованных в использовании разработанной системы, — и временных затрат на автоматизированную обработку данных был проведен сравнительный анализ эффективности разработанной информационной системы, согласно которому для обработки территории,

соответствующей одному муниципальному району Республики Башкортостан, требуется 62,5 мин для автоматизированной обработки данных, против 107,5 мин при ручной обработке.

Достигнута экономия времени порядка 40 % (рис. 6). При экстраполяции на всю территорию Республики Башкортостан экономия времени достигает 47 ч.

Другим преимуществом системы являются более низкие требования к квалификации специалиста, осуществляющего анализ, т. к. экспертные знания, необходимые для обработки, уже формализованы.

Размещение объектов техногенной опасности имеет важное практическое значение, связанное с необходимостью контроля над экологическими рисками и возможными последствиями техногенного характера. Разработанная информационная система объектов техногенной опасности в условиях неопределенности позволяет осуществлять размещение объектов техногенной опасности по заданным критериям, обеспе-



Рис. 6. Анализ эффективности информационной системы объектов техногенной опасности в условиях неопределенности

чивая формализацию экспертных знаний и позволяя сократить временные затраты на проводимый анализ.

В настоящий момент функциональность системы расширяется в части возможности динамического изменения критериев анализа и пороговых значений термов функции принадлежности. Внедрение дополнительного функционала позволит использовать информационную систему для анализа территории в любых целях, не только для размещения полигонов отходов, и позволит учитывать специфические особенности территории, которая будет подвергнута анализу. К примеру, для гористой местности возможно добавление такого критерия, как частота и сила оползней по шкале Медведева–Шпонхойера–Карника.

Однако дальнейшее использование системы возможно и без расширения ее функциональности. Существующая модель обработки данных пригодна для анализа всей территории Российской Федерации или других стран, а выигрыш по времени при анализе участков территории таких масштабов с помощью информационной системы по сравнению с ручной обработкой информации возрастает пропорционально площади исследуемой территории.

Результаты исследований, представленные в статье, частично поддержаны грантом № 15-08-01758–А «Методологические и методические основы анализа техногенной безопасности в условиях неопределенности состояния объектов управления».

СПИСОК ЛИТЕРАТУРЫ

1. Федеральная служба государственной статистики. Образование отходов производства и потребления по видам экономической деятельности [электронный ресурс]// URL: http://www.gks.ru/free_doc_new_site/oxrana/tab1/oxr_otxod1.xls/ (дата обращения: 19.12.2014).

xls/ (дата обращения: 19.12.2014).

2. Zadeh L.A. Fuzzy sets // Information and Control. 1965. Vol. 8. No. 3. Pp. 338–353.

3. Klaua D., Gottwald S. An early approach toward graded identity and graded membership in set

theory // *Fuzzy Sets and Systems*. Vol. 161, No. 18. Pp. 114–119.

4. **Давлетбакова З.Л.** Применение методов нечеткой логики в пространственном анализе объектов размещения отходов // *Актуальные проблемы науки и техники. Информационные и инфокоммуникационные технологии, естественные науки: сб. трудов*. Уфа: Изд-во УГАТУ, 2012. Т. 1. С. 303–306.

5. **Алтунин А.Е., Семухин М.В.** Модели и алгоритмы принятия решений в нечетких условиях: Монография. Тюмень: Изд-во Тюменского гос. ун-та, 2000. 352 с.

6. Санитарно-защитные зоны и санитарная классификация предприятий, сооружений и иных объектов // СанПиН 2.2.1/2.1.1.1200-03. Зарег. в Мин-ве юстиции РФ 29.04.2003. № 4459.

7. **Христодуло О.И., Салимзянов И.Ф., Гареева Н.Р.** Многокритериальная математическая модель позиционирования объектов техногенной опасности на основе нечеткой логики //

Геоинформационные технологии в проектировании и создании корпоративных информационных систем: межвуз. науч. сб. Уфа: Изд-во УГАТУ, 2015. С. 55–64.

8. **Заде Л.** Понятие лингвистической переменной и его применение к принятию приближенных решений. М.: Мир, 1976. 166 с.

9. **Штовба С.Д.** Введение в теорию нечетких множеств и нечеткую логику. Винница: Континент-Прим, 2003. 198 с.

10. **Зак Ю.А.** Принятие решений в условиях нечетких и размытых данных: Fuzzy-технологии. М.: Книжный дом «ЛИБРОКОМ», 2013. 352 с.

11. **Павлов С.В., Давлетбакова З.Л., Абдуллин А.Х.** Информационная система поддержки принятия решений по управлению отходами на территории Республики Башкортостан на основе базы нечетких знаний // *Электротехнические и информационные комплексы и системы*. 2013. № 4. Т. 9. С. 113–120.

REFERENCES

1. *Federalnaya sluzhba gosudarstvennoy statistiki. Obrazovaniye otkhodov proizvodstva i potrebleniya po vidam ekonomicheskoy deyatel'nosti [Federal State Statistics Service. Waste production and consumption by economic activity]*. Available: http://www.gks.ru/free_doc_new_site/oxrana/tab1/oxr_otxod1.xls/ (Accessed: 19.12.2014). (rus)

2. **Zadeh L.A.** Fuzzy sets. *Information and Control*, 1965, Vol. 8, No. 3, Pp. 338–353.

3. **Klaua D., Gottwald S.** An early approach toward graded identity and graded membership in set theory. *Fuzzy Sets and Systems*, Vol. 161, No. 18, Pp. 114–119.

4. **Davletbakova Z.L.** Primeneniye metodov nechetkoy logiki v prostranstvennom analize obyektov razmeshcheniya otkhodov [Application of fuzzy logic in the spatial analysis of waste disposal facilities]. *Aktualnyye problemy nauki i tekhniki: Informatsionnyye i infokommunikatsionnyye tekhnologii, yestestvennyye nauki [Actual problems of science and technology and information and communication technology Information, Science]*, Ufa: UGATU Publ., 2012, Vol. 1, Pp. 303–306. (rus)

5. **Altunin A.Ye., Semukhin M.V.** Modeli i algoritmy prinyatiya resheniy v nechetkikh usloviyakh [Models and algorithms for decision making in fuzzy conditions]. Tyumen: Tyumenskii gos. Universitet Publ., 2000, 352 p. (rus)

6. *Sanitarno-zashchitnyye zony i sanitarnaya klassifikatsiya predpriyatiy, sooruzheniy i inykh obyektov [Sanitary protection zones and sanitary classification*

of enterprises, buildings and other facilities]. SanPiN 2.2.1/2.1.1.1200-03, zareg. v Ministerstve yustitsii RF 29.04.2003, No. 4459. (rus)

7. **Khristodulo O.I., Salimzyanov I.F., Gareyeva N.R.** mnogokriterialnaya matematicheskaya model pozitsionirovaniya obyektov tekhnogennoy opasnosti na osnove nechetkoy logiki [Multi-criteria mathematical model of positioning objects man-made hazards on the basis of fuzzy logic]. *Geoinformatsionnyye tekhnologii v proyektirovani i sozdani korporativnykh informatsionnykh sistem [Geoinformation technologies in the design and construction of corporate information systems]*, Ufa: UGATU Publ., 2015, Pp. 55–64. (rus)

8. **Zade L.** *Ponyatiye lingvisticheskoy peremennoy i yego primeneniye k prinyatiyu priblizhennykh resheniy [The concept of linguistic variable and its application to decision-making close]*, Moscow: Mir Publ., 1976, 166 p. (rus)

9. **Shtovba S.D.** *Vvedeniye v teoriyu nechetkikh mnozhestv i nechetkuyu logiku [Introduction to the theory of fuzzy sets and fuzzy logic]*, Vinnitsa: Kontinent-Prim Publ., 2003, 198 p. (rus)

10. **Zak Yu.A.** *Prinyatiye resheniy v usloviyakh nechetkikh i razmytykh dannykh: Fuzzy-tekhnologii [Decision making under fuzzy and fuzzy data: Fuzzy-technology]*, Moscow: Knizhnyy dom «LIBROKOM» Publ., 2013, 352 p. (rus)

11. **Pavlov S.V., Davletbakova Z.L., Abdullin A.Kh.** Informatsionnaya sistema podderzhki prinyatiya resheniy po upravleniyu otkhodami na territorii Respubliki Bashkortostan na osnove bazy

nchetkikh znaniy [Information Decision Support System for Waste Management in the Republic of Bashkortostan on the basis of fuzzy knowledge base].

Elektrotekhnicheskiye i informatsionnyye komplekсы i sistemy [Electrical and information systems and systems], 2013, No. 4, Vol. 9, Pp. 113–120. (rus)

ХРИСТОДУЛО Ольга Игоревна – *заведующий кафедрой геоинформационных систем Уфимского государственного авиационного технического университета, доктор технических наук.*

450000, Россия, Республика Башкортостан, г. Уфа, ул. К. Маркса, д. 12.

E-mail: o-hristodulo@mail.ru

KHRISTODULO Olga I. *Ufa State Aviation Technical University.*

450000, Karl Marx Str. 12, Ufa, Republic of Bashkortostan, Russia.

E-mail: o-hristodulo@mail.ru

САЛИМЗЯНОВ Ильяс Фаритович – *аспирант кафедры геоинформационных систем Уфимского государственного авиационного технического университета.*

450000, Россия, Республика Башкортостан, г. Уфа, ул. К. Маркса, д. 12.

E-mail: mau1er@yandex.ru

SALIMZYNOV Iiyas F. *Ufa State Aviation Technical University.*

450000, Karl Marx Str. 12, Ufa, Republic of Bashkortostan, Russia.

E-mail: mau1er@yandex.ru

ГАРЕЕВА Наиля Рифовна – *студентка кафедры геоинформационных систем Уфимского государственного авиационного технического университета.*

450000, Россия, Республика Башкортостан, г. Уфа, ул. К. Маркса, д. 12.

E-mail: nailyarg@yandex.ru

GAREEVA Nailya R. *Ufa State Aviation Technical University.*

450000, Karl Marx Str. 12, Ufa, Republic of Bashkortostan, Russia.

E-mail: nailyarg@yandex.ru



DOI: 10.5862/JCSTCS.229.6

УДК 681.51:621:391

Г.Ф. Малыхина, И.А. Кислицына

ИЗМЕРЕНИЕ ПАРАМЕТРОВ ДВИЖЕНИЯ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

G.F. Malykhina, I.A. Kislitcina

MEASUREMENT OF MOTION CHARACTERISTICS USING NEURAL NETWORKS

Предложен нейросетевой алгоритм измерения характеристик движения спускаемого аппарата на лунную поверхность, использующий фотонный преобразователь. Измерительная система основана на регистрации рассеянного излучения фотонов и включает источник и четыре детектора фотонов. Модель предназначена для использования на начальных этапах проектирования системы измерения текущей высоты, скорости движения и наклона спускаемого аппарата относительно подстилающей поверхности. Модель позволяет выполнить анализ геометрии расположения источников и детекторов излучения, оценить среднее значение интенсивности потока фотонов и исследовать влияние на результаты измерения состава подстилающей поверхности. Алгоритм реализует нелинейную модель пространства состояний. Обучение алгоритма выполняется по данным компьютерного моделирования. Приведены результаты моделирования работы алгоритма и выполнена оценка погрешности измерения высоты, скорости и угла наклона спускаемого аппарата.

НЕЙРОСЕТЕВОЙ АЛГОРИТМ; ИЗМЕРЕНИЕ; ДВИЖЕНИЕ СПУСКАЕМОГО АППАРАТА.

A neural-network-based algorithm is offered for measuring the motion characteristics of a landing module. The measurement system is based on recording photon scattering. It uses a central photon source and four photon detectors for measuring the altitude, the traverse speed and the slope angle of the landing module. The model is intended for the initial stage of developing the information measuring system. The model allows to estimate the results of altitude, traverse speed and slope angle measurement, analyze the geometrical arrangement of photon sources and photon detectors, investigate the influence of chemical composition of underlying surface. The algorithm uses the information obtained from the photon detectors arranged on the lunar landing module and implements a state-space model of the landing module. A computer model of module motion produces training data. The generalization results allow to estimate measurement uncertainty. The simulation results of the algorithm and the estimated error of measurement of height, speed and angle of descent vehicle are presented.

NEURAL NETWORK ALGORITHM; LENDER MODULE MOTION; MEASUREMENT.

Для успешной посадки спускаемого аппарата на поверхность Луны необходимо выполнять достаточно точные измерения параметров движения, текущей высоты, скорости и угла наклона спускаемого аппарата над подстилающей поверхностью в реальном времени. Результаты измерения предназначены для управления двигателями мягкой посадки. Измерения выполняются в условиях неравномерного спуска, обусловленного отсутствием атмосферы,

что исключает спуск на парашюте, наличием масконов, приводящих к неравномерности силы тяжести [1].

Особенностью измерения является неточное представление об элементном составе подстилающей поверхности в месте посадки. Наличие слоя пыли на поверхности делает измерение нечетким. Достаточно полный натурный эксперимент невозможно провести в земных условиях, поэтому повышается роль моделирования.

Существуют высотомеры различных диапазонов частот: радиоволнового ($3 \cdot 10^9$ – $30 \cdot 10^9$ Гц), инфракрасного ($1,5 \cdot 10^{12}$ – $400 \cdot 10^{12}$ Гц), видеокамеры ($270 \cdot 10^{12}$ – $790 \cdot 10^{12}$ Гц), рентгеновского ($3 \cdot 10^{16}$ – $3 \cdot 10^{19}$ Гц), фотонного ($> 3 \cdot 10^{19}$ Гц).

Лазерный дальномер – стандартное измерительное средство для автоматических космических кораблей. В качестве примера рассмотрим следующие проекты, приведенные в табл. 1: VepiColombo (совместная для ЕКА и JAXA космическая автоматическая миссия к Меркурию), Lunar Orbiter Laser Altimeter (программа NASA Lunar Reconnaissance Orbiter по картографированию поверхности Луны), Mercury Laser Altimeter (программа NASA Messenger по картографированию поверхности Меркурия), «Фобос-Грунт» (посадка космического аппарата (КА) на поверхность спутника Марса – Фобоса, забор образцов грунта и доставка их на Землю) [2, 3].

Рассмотренные в табл. 1 высотомеры применяются на больших расстояниях. Для определения малых высот при осуществлении посадки летательных аппаратов используют лазерные дальномеры-высотомеры, указанные в табл. 2. Однако и эти высотомеры имеют недостатки. В-первых, для обеспечения высокой надеж-

ности в условиях длительной космической экспедиции (вакуум, различные виды радиации, перепады температуры, вибрация на этапе вывода в космос и пр.) существенно осложняется задача построения системы с учетом минимизации массы и энергопотребления. Во-вторых, лазерное излучение является остронаправленным, что повышает чувствительность системы к случайным препятствиям и облаку пыли, образованному в результате работы двигателей мягкой посадки.

Для определения малых высот спускаемого аппарата над подстилающей поверхностью может использоваться радиолокационный способ измерений. С помощью радиовысотомеров была успешно осуществлена мягкая посадка целого ряда советских автоматических межпланетных станций «Луна» (от «Луна-16» до «Луна-24») для изучения естественного спутника Земли. Существующие в настоящее время модели высотомеров представлены в табл. 2. Бортовые фазовые измерители дальности (высоты) обладают выраженным преимуществом по сравнению с высотомерами малых высот, использующими импульсный или частотно-модулированный зондирующий сигнал, поскольку позволяют измерять сверхмалые высоты вплоть до единиц

Таблица 1

Лазерные дальномеры, используемые для управления космическим аппаратом

КА-носитель	Clementine	NEAR Shoemaker	Mars Global Surveyor	MESSENGER	VepiColombo	LRO	Фобос-Грунт
Планета назначения	Луна	Эрос	Марс	Меркурий	Меркурий	Луна	Фобос
Дата запуска	25 января 1994	17 февраля 1996	7 ноября 1996	3 августа 2004	2017	19 июня 2009	2011
Название прибора	LIDAR	NLR	MOLA	MLA	BELA	LOLA	ЛВ-В
Диапазон дальностей, км	400...8300	35...50	400	200...15 000	400...1500	50	0,5...100

Таблица 2

Лазерные дальномеры-высотомеры, используемые для посадки летательных аппаратов

Наименование прибора	Диапазон высот, м	Погрешности измерений, м
ДЛ-1 [4]	2...600	Среднеквадратическое отклонение значений измеренной дальности не более 0,2. Систематическое отклонение значений измеренной дальности не более $\pm (0,2 + 0,001 \times D)$
ДЛ-5М [4]		Максимальная ошибка измерения - в диапазоне от 2 до 200 м: 0,5; - диапазоне от 200 до 1000 м: 2,2
ДЛ-8 [4]	0,3...150	Систематическая погрешность измерения $\pm(0,03+0,001 \times H)$. Среднеквадратическая погрешность измерения не более 0,03

дециметров. Использование же в фазовом дальномере нескольких частотных шкал (например, трех) дает возможность измерять дальности до нескольких километров, т. е. перекрывать четыре декады. В отечественной и зарубежной литературе указывается на возможность реализации бортовых многолучевых систем доплеровского измерения скорости (ДИС) КА и измерителей наклонной дальности (высоты) до зондируемой поверхности, использующих фазовый метод. Такие системы осуществляют, кроме того, построение радиовертикали, поэтому измеритель, входящий в их состав, назван ниже *радиовертикантом*. [5]. С помощью системы ДИС могут быть успешно реализованы требования к точности измерения высоты $\pm 0,3 + 2 \% H$.

Как видно из табл. 3, точность радиовысотомеров уменьшается обратно пропорционально высоте. Кроме того, радиовысотомеры не могут работать через обшивку спускаемого аппарата и через плазму работающего двигателя мягкой посадки.

В 1966–1968 гг. в ЦНИИ Робототехники и технической кибернетики был разработан высотомер типа «Кактус». [9] Принцип действия высотомера основан на регистрации обратно рассеянного от подстилающей поверхности гамма-излучения. Сигнал, отраженный от подстилающей поверхности и регистрируемый детектором, несет информацию о высоте, которая отсчитывается от уровня поверхности.

Система выполняет непосредственное измерение высоты и косвенное измерение скорости и угла наклона исходя из результатов измерения высоты. Для измерения высоты, превышающей 10 м, целесообразно использовать радиовысотомер, но на меньших высотах радиовысотомер дает значительные погрешности измерения методического характера, что приводит к необходимости разработки высотомера, основанного на регистрации рассеянного гамма-излучения. Преимуществами фотонного высотомера являются:

- способность выполнять измерения через обшивку спускаемого аппарата в условиях плазмы двигателей мягкой посадки;

- нечувствительность системы к слою пыли на поверхности планеты и к случайным препятствиям;

- обеспечение высокой точности измерений на малых высотах.

Фотонный высотомер характеризуется случайными некоррелированными шумами, распределенными по нормальному закону пуассоновского типа. Относительная погрешность измерения фотонного высотомера уменьшается при измерении на малых высотах, что является положительным свойством для управления двигателем посадки. Поэтому в системе посадки целесообразно использовать совместно радиовысотомер на высотах, превышающих 10 м, и фотонный высотомер на высотах от 0,3 до 10 м.

Таблица 3

Бортовые радиовысотомеры

Наименование прибора	Диапазон высот, м	Погрешности измерений, м
А-035 (РВ-21) [6]	0...1500; 0...6000; 0...11000	от 0 до 20 м (+ 2); более 20 м (+ 0,1 Н)
А-037 [7]	0...750	в диапазоне углов крена и тангажа $\pm 40^\circ$: при $0 \leq H < 60$ м: $\pm(0,6+0,05 Н)$ при $60 \leq H \leq 750$ м: $\pm 0,06 Н$
А-040 [7]	10...1000	$\pm 0,05 Н$
А-040-01 [8]	10...1000	в диапазоне высот от 10 до 75 м $\pm (1 + 0,08 Н)$ в диапазоне высот от 75 м $\pm 0,08 Н$
А-052 [7]	0...1500	по цифровому выходу: $\pm(0,45 + 0,01 Н)$ по индикатору: $\pm 2,0$ или $\pm 0,1 Н$
А-053 [7]	0...1500	по цифровому выходу: $\pm 0,45$ или $\pm 0,02 Н^*$ по индикатору: $\pm(1,5 + 0,05 Н)$
А-075 [7]	0...25000	$(1,5 + 0, 015 Н)$
А-076 [7]	0...20000	$\pm(1 + 0, 01 Н)^*$, * в диапазоне Н от 20 до 10000 м
А-078 [8]	0...10000	$\pm(0,5 + 0,01 Н)$
А-098 [7]	0...8000	$\pm(1,5+0,015 Н)$
РВ-2	0...120,100...1200	$\pm 5 \dots \pm 10$
РВ-5М (А-037) [6]	0...750	от 0 до 10 м: $\pm 0,8$ от 10 до 750 м: $\pm 8 \%$
РВ-10	80...12000	15-150 м: $\pm 0,25$
РВ-17	100...1700	15 м: $0,25 Н$

Для обработки результатов измерения целесообразно использовать нейросетевые алгоритмы, преимуществом которых является фильтрация в процессе измерения шума фотонного преобразователя, способность обучения системы измерения по результатам имитационного эксперимента, способность обобщения результатов обучения на других данных, которые не были предъявлены в процессе обучения, возможность дообучения в процессе получения новых данных в реальных экспериментах.

Задачей настоящей статьи является разработка нейросетевого алгоритма измерения параметров движения спускаемого аппарата на лунную поверхность на основе данных компьютерного моделирования.

В основу моделирования положено соотношение Клейна–Нишины–Тамма, характеризующее взаимодействие фотонов с веществом.

Условия посадки на Луну серьезно отличаются от земных условий отсутствием атмосферы, неравномерным движением, высоким фоном радиационного излучения, составом грунта.

Геометрия системы измерения параметров движения. Система имеет источник фотонного излучения (ФИ), размещенный в центре спускаемого аппарата, и четыре детектора (Д1–Д4), расположенные по краям на равном расстоянии l от источника. Спускаемый аппарат может быть наклонен относительно подстилающей поверхности,

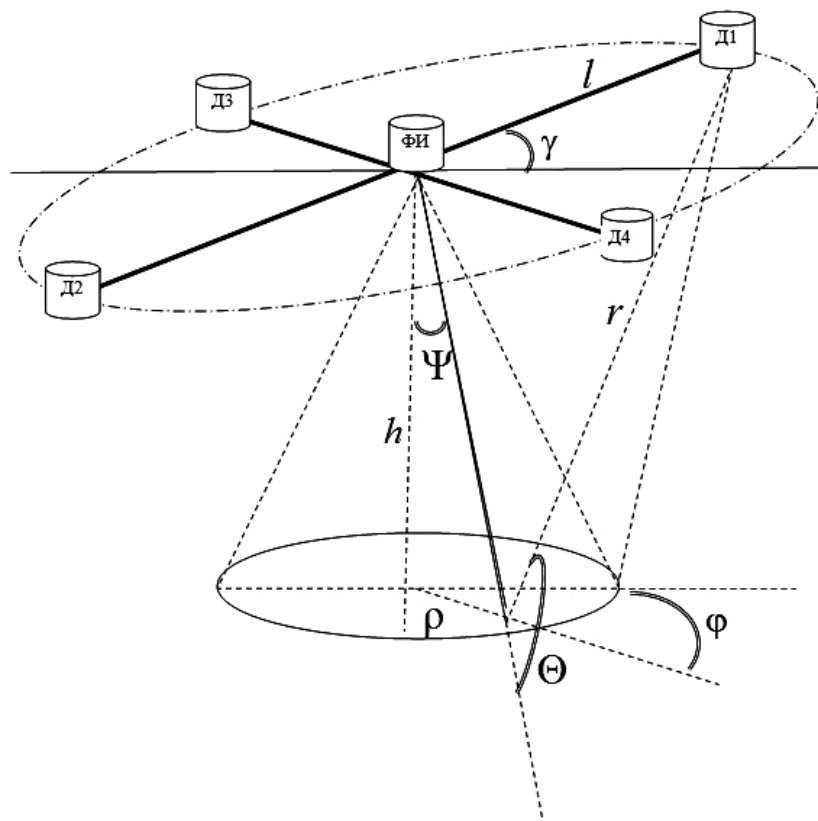


Рис. 1. Расположение источника и приемников излучения

причем угол γ_1 характеризует наклон оси, на которой расположены приемники Д1 и Д2, угол γ_2 – наклон оси приемников Д3-Д4. На рис. 1 представлена геометрия системы, где h – текущее значение высоты спускаемого аппарата, r – расстояние от рассеивающего элемента до детектора, ρ – расстояние между проекцией центра ФИ и элементом рассеяния, Ψ – угол падения прямого потока фотонов, $depth$ – глубина, на которой происходит рассеивание, $\Psi \leq \Psi_{max}$, где Ψ_{max} – угол коллимации, φ – угол между проекцией оси детекторов Д1-Д2 и радиусом, проходящим через элемент рассеяния фотонов, Θ – угол рассеяния фотонов. Приемник излучения имеет конечные размеры, поэтому диапазон углов рассеяния фотонов, достигающих детектор Д1, составляет $[\Theta_1, \Theta_2]$.

Моделирование фотонного преобразователя. Интенсивность излучения, воспринимаемого одним из детекторов, например Д1, зависит от угла наклона γ_1 оси детек-

тора над подстилающей поверхностью, высоты h , расстояния r между источником и детектором излучения, угла рассеяния Θ , отношения α энергии фотона к энергии покоящегося электрона $\alpha \approx 1,129$, сечения комптоновского рассеяния на один электрон, отнесенного к единице телесного угла σ_0 . Интенсивность рассеянного под углом Θ фотонного излучения, зависящая от геометрии расположения источника и приемников излучения, может быть вычислена по формуле:

$$I_{\Theta}(r, l, \gamma, \Psi, h, \varphi, \Theta) = I_0 \frac{3\sigma_0}{16\pi r^2} \cdot \frac{(1 + \cos^2(\Theta))}{[1 + \alpha(1 - \cos^2(\Theta))]^2} \times \left\{ 1 + \frac{\alpha^2(1 - \cos^2(\Theta))^2}{[1 + \alpha(1 - \cos^2(\Theta))] \cdot (1 + \cos^2(\Theta))} \right\}. \quad (1)$$

Интенсивность потока регистрируемых гамма-квантов может быть получена интегрированием по объему рассеивающего вещества:

$$I_{\text{detect}} = \int_{\text{depth}} \int_{\varphi} \int_{\rho} e^{-\mu(E) \cdot \text{depth}} I_{\Theta}(r, l, \gamma, \Psi, h, \varphi, \Theta) \times d(\text{depth}) d\varphi d\rho, \quad (2)$$

где $\mu(E)$ – коэффициент массового поглощения, зависящий от энергии фотонов; ρ – плотность слоя рассеивающей поверхности; depth – глубина, на которой расположен рассеивающий элемент.

Соотношение (2) характеризует среднюю интенсивность регистрируемого потока фотонов. Текущее значение интенсивности содержит шум, некоррелированный и распределенный по закону Пуассона. При достаточно большой интенсивности потока шум имеет Гауссово распределение, тогда интенсивность потока фотонов представляет собой сумму систематической и случайной составляющей:

$$I_{\text{detect}}(n) = I_{\text{detect}}(n) + N(n), \\ \sigma(N(n)) = \sqrt{I_{\text{detect}}(n)}.$$

Моделирование интенсивности регистрируемых потоков при различных высотах, углах наклона и скоростях спуска позволяет выполнять проектирование системы посадки.

Моделью объекта и системы измерения является нелинейная динамическая система, представленная в терминах пространства состояний. Измерительная система работает в условиях неполной информации об объекте измерения и об условиях измерения. Динамическое поведение системы характеризуется уравнением управления

$$\mathbf{x}(n+1) = \Phi(\mathbf{W}_A \mathbf{x}(n) + \mathbf{W}_B \mathbf{u}(n)) \quad (4)$$

и уравнением наблюдения

$$\mathbf{y}(n) = \Phi(\mathbf{W}_C \mathbf{x}(n) + \mathbf{e}(n)), \quad (5)$$

где $\mathbf{x}(n)$ – $q \times 1$ -вектор, характеризующий состояние нелинейной динамической системы дискретного времени; $\mathbf{u}(n)$ – $m \times 1$ входной вектор управляющих воздействий; $\mathbf{y}(n)$ – $p \times 1$ вектор, характеризующий выход системы; $\mathbf{e}(n)$ – вектор погрешностей измерения; \mathbf{W}_A – матрица размером $q \times q$; \mathbf{W}_B – матрица размером $q \times (m+1)$; \mathbf{C} – матрица размером $p \times q$; $\Phi()$ – функция, характеризующая нелинейность. Про-

странства R^m , R^q и R^p называются входным пространством, пространством состояний и выходным пространством соответственно. Размерность пространства состояний q определяет порядок системы. Таким образом, модель пространства состояний характеризует нелинейную динамическую систему дискретного времени порядка q с m входами и p выходами.

Погрешность, вызванная шумом радиоизотопного преобразователя, не является аддитивной, случайная составляющая погрешности зависит от интенсивности потока фотонов.

Рассмотрим уравнение системы с одним входом и одним выходом:

$$\mathbf{x}(n+1) = \Phi(\mathbf{W}_A \mathbf{x}(n) + \mathbf{W}_B u(n)), \quad (6)$$

$$y(n) = \Phi(\mathbf{W}_C \mathbf{x}(n) + e(n)). \quad (7)$$

Линеаризация уравнения (6) в точке $\mathbf{x} = \mathbf{0}$, $u = 0$ позволяет получить уравнение:

$$\delta \mathbf{x}(n+1) = \Phi'(0) \mathbf{W}_A \delta \mathbf{x}(n) + \Phi'(0) \mathbf{W}_B \delta u(n).$$

Введя обозначение $\mathbf{A} = \Phi'(0) \mathbf{W}_A$, $\mathbf{B} = \Phi'(0) \mathbf{W}_B$, получим рекуррентные соотношения для моделирования динамической системы:

$$\delta \mathbf{x}(n+1) = \mathbf{A} \delta \mathbf{x}(n) + \mathbf{B} \delta u(n)$$

$$\delta \mathbf{x}(n+2) = \mathbf{A}^2 \delta \mathbf{x}(n) + \mathbf{A} \mathbf{B} \delta u(n) + \mathbf{B} \delta u(n+1)$$

...

$$\delta \mathbf{x}(n+q) = \mathbf{A}^q \delta \mathbf{x}(n) + \mathbf{A}^{q-1} \mathbf{B} \delta u(n) + \mathbf{A}^{q-2} \mathbf{B}^2 \delta u(n+1) + \dots + \mathbf{A} \mathbf{B} \delta u(n+q-2) + \mathbf{B} \delta u(n+q-1).$$

Вектор входных воздействий имеет размерность q :

$$u_p(n) = [u(n), u(n+1), \dots, u(n+q-1)]^T.$$

Уравнение измерения характеристик объекта после линеаризации уравнения (7) и применения обозначения $\mathbf{C} = \Phi'(0) \mathbf{W}_C$ приобретает вид:

$$\delta y(n) = \mathbf{C}^T \delta \mathbf{x}(n) + \delta e(n)$$

$$\delta y(n+1) = \mathbf{C}^T \mathbf{A} \delta \mathbf{x}(n) + \mathbf{C}^T \mathbf{B} \delta u(n) + \delta e(n+1)$$

$$\delta y(n+2) = \mathbf{C}^T \mathbf{A}^2 \delta \mathbf{x}(n) + \mathbf{C}^T \mathbf{A} \mathbf{B} \delta u(n) + \mathbf{C}^T \mathbf{B} \delta u(n+1) + \delta e(n+2)$$

...

$$\begin{aligned} & \dots \\ \delta y(n+q-1) &= \mathbf{C}^T \mathbf{A}^{q-1} \delta \mathbf{x}(n) + \\ &+ \mathbf{C}^T \mathbf{A}^{q-2} \mathbf{B} \delta u(n) + \dots + \mathbf{C}^T \mathbf{A} \mathbf{B} \delta u(n+q-3) + \\ &+ \mathbf{C}^T \mathbf{B} \delta u(n+q-2) + \delta e(n+q-1). \end{aligned}$$

Нелинейная авторегрессионная сеть с внешними входами:

$$\begin{aligned} \delta y(n) + \delta y(n+1) + \dots + \delta y(n+q) &= \\ &= (\mathbf{C}^T + \mathbf{C}^T \mathbf{A} + \dots + \mathbf{C}^T \mathbf{A}^{q-1}) \delta \mathbf{x}(n) + \\ &+ (\mathbf{C}^T \mathbf{B} + \mathbf{C}^T \mathbf{A} \mathbf{B} + \dots + \mathbf{C}^T \mathbf{A}^{q-1} \mathbf{B}) \delta u(n) + \\ &+ (\mathbf{C}^T \mathbf{B} + \mathbf{C}^T \mathbf{A} \mathbf{B} + \dots + \mathbf{C}^T \mathbf{A}^{q-2} \mathbf{B}) \delta u(n+1) + \dots \\ &+ \mathbf{C}^T \mathbf{B} \delta u(n+q) + \delta e(n) + \delta e(n+1) + \dots \\ &+ \delta e(n+q-1). \end{aligned}$$

Заменим n на $n-q+2$ и запишем уравнение измерения для линеаризованной системы:

$$\begin{aligned} y(n+1) &= -y(n) - y(n-1) - \dots \\ &- y(n-q+2) + (\mathbf{C}^T \mathbf{B} + \mathbf{C}^T \mathbf{A} \mathbf{B} + \dots \\ &+ \mathbf{C}^T \mathbf{A}^{q-1} \mathbf{B}) \mathbf{x}(n-q+2) + (\mathbf{C}^T \mathbf{B} + \mathbf{C}^T \mathbf{A} \mathbf{B} + \dots \\ &+ \mathbf{C}^T \mathbf{A}^{q-2} \mathbf{B}) u(n) + \dots + \mathbf{C}^T \mathbf{B} u(n+q-2) + \\ &+ \delta e(n-q+1) + \delta e(n-q) + \dots + \delta e(n). \end{aligned} \quad (8)$$

Матрица $\mathbf{M} = [\mathbf{C}, \mathbf{C} \mathbf{A}^T, \dots, \mathbf{C} (\mathbf{A}^T)^{q-1}]$ является матрицей наблюдаемости линеаризованной системы.

Запишем в виде суммы:

$$\begin{aligned} y(n) &= - \sum_{k=n-q+2}^n y(k) + \\ &+ \sum_{k=0}^{q-1} \mathbf{C}^T \mathbf{A}^k \mathbf{B} \cdot \mathbf{x}(n-q+2) + \sum_{k=0}^{q-2} \mathbf{C}^T \mathbf{A}^k \mathbf{B} \cdot u(n) + \dots \\ &+ \mathbf{C}^T \mathbf{B} u(n-q+1) + \sum_{k=n-q+1}^n e(k). \end{aligned}$$

Уравнение измерения (8) реализует рекуррентная нейронная сеть с одним входом и одним выходом. Увеличение числа входов и числа выходов рекуррентной нейронной сети позволит выполнять измерение всех параметров движения спускаемого аппарата. Вектор входных параметров содержит пять компонент, которые представляют число фотонов, регистрируемых на заданном временном интервале каждым из четырех детекторов $I_i = \left(\frac{I(n)}{I_0} \right)_{D_i}$, $i = 1, 2, 3, 4$ и значения высоты h_{radio} , измеренной

радиовысотомером:

$$\mathbf{u}(n) = \left\{ \left(\frac{I_1(n)}{I_0} \right)_{D1}, \left(\frac{I_2(n)}{I_0} \right)_{D2}, \left(\frac{I_3(n)}{I_0} \right)_{D3}, \left(\frac{I_4(n)}{I_0} \right)_{D4}, h_{\text{radio}} \right\}.$$

Вектор измеренных сигналов содержит четыре компоненты: текущую высоту, скорость и два угла наклона осей детекторов Д1-Д2 и Д3-Д4.

$$\mathbf{y}(n) = \{h(n), v(n), \gamma_1(n), \gamma_2(n)\}.$$

Модель измерения представляет собой нелинейную авторегрессию с внешними входами. Предложено использовать полносвязную нейронную сеть с одним внутренним слоем с логистической активационной функцией и с линейным выходным слоем. Измеренное значение сигнала

$$y_k(n) = \sum_{i=0}^{L1} \varphi_k \left(w_{ik} \left(\sum_{p=0}^P \sum_{j=0}^{L0} \varphi_j(w_{ij}(p) \times x_j(n-p)) + b_i \right) \right) + b_k,$$

$$i = 0..L1, p = 0..P, j = 0..L0, k = 0..K,$$

где $L0$ – число элементов входного слоя; $L1$ – число нейронов скрытого слоя; P – порядок системы; $w_{ij}(p), b_i$ – веса и смещения внутреннего слоя нейронов; w_{ik}, b_k – веса и смещения выходного слоя нейронов. На вход нейронной сети поступают текущие $p = 0$ и задержанные на $p = 1, 2$ значения интенсивности потока фотонов, регистрируемых детекторами, высота по радиовысотомеру:

$$\begin{aligned} \mathbf{x}_1(n) &= [I_1(n), I_2(n), I_3(n), I_4(n), I_1(n-1), \\ &I_2(n-1), I_3(n-1), I_4(n-1), I_1(n-2), \\ &I_2(n-2), I_3(n-2), I_4(n), h_{\text{radio}}(0)]^T \end{aligned}$$

и задержанные на $p = 1, 2, 3$ предыдущие значения высоты, углов наклона и скорости:

$$\begin{aligned} \mathbf{x}_2(n) &= [h(n-1), h(n-2), h(n-3), v(n-1), \\ &v(n-2), v(n-3), \gamma_1(n-1), \gamma_1(n-1), \gamma_1(n-1), \\ &\gamma_2(n-1), \gamma_2(n-1), \gamma_2(n-1)]^T. \end{aligned}$$

Структура рекуррентной нейронной сети с задержками входного сигнала по-

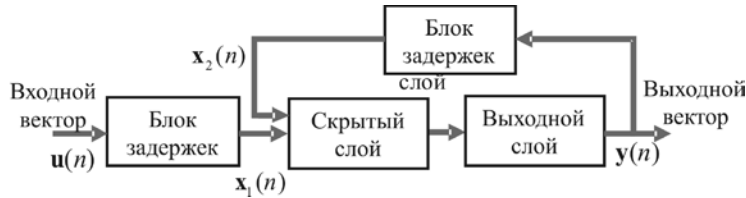


Рис. 2. Рекуррентная нейронная сеть с задержками входного сигнала

казана на рис. 2. Для обучения нейронной сети применен метод Бройдена–Флетчера–Гольдфарба–Шанно, который относится к группе квазиньютоновских методов оптимизации и считается лучшим из них. Вычислительная сложность этого метода более высокая, чем у градиентных методов. Однако обучение нейронной сети выполняется на этапе градуировки системы, поэтому длительность обучения не оказывает существенного влияния на работу измерительной системы. Оптимальный размер нейронной сети, содержащей три нейрона скрытого слоя и четыре нейрона выходного слоя, был подобран путем постепенного наращивания числа слоев и числа нейронов в слое и сопоставления среднеквадратической ошибки обучения и обобщения. Это позволило избежать эффекта переобучения нейронной сети.

Вектор входных параметров $x_1(n)$ нормализован. Нормализация включает следующие преобразования:

приведение всех компонент входных векторов к виду с нулевым средним значением и единичной дисперсией;

вычисление главных компонент векторов.

Применение главных компонент позволяет на 30 % уменьшить объем обучающей выборки. В процессе моделирования возможно включение также данных физического эксперимента, выполняемого в земных условиях [10].

Результаты моделирования. Для моделирования системы измерения параметров движения спускаемого аппарата разработана программа, включающая моделирование:

фотонного преобразователя при различном расположении источника и приемников излучения;

шума фотонного преобразователя;

нейросетевого алгоритма измерения параметров движения.

С использованием этой программы вы-

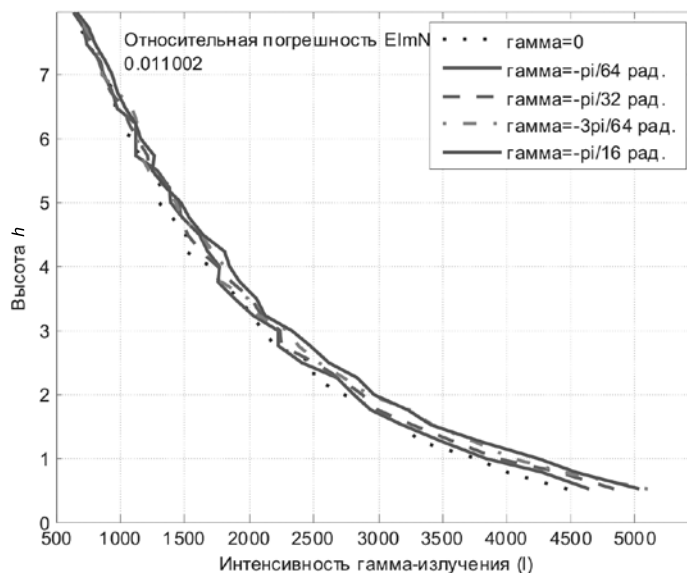


Рис. 3. Зависимость измеренного значения высоты от количества зарегистрированных фотонов

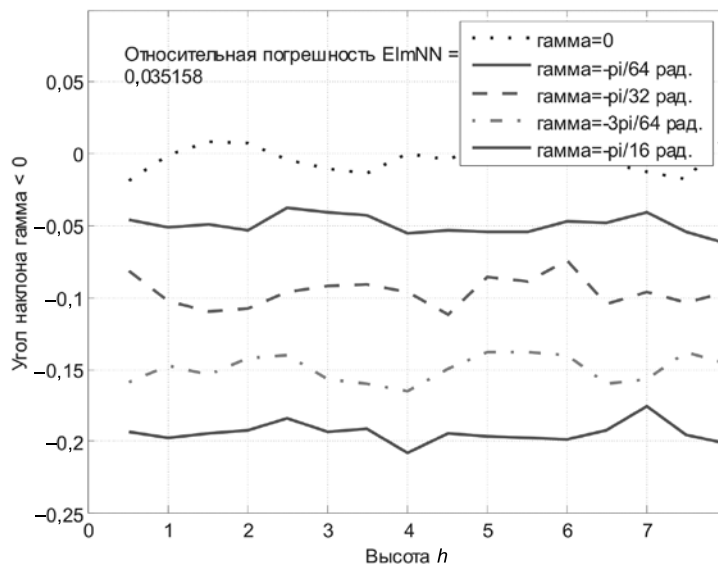


Рис. 4. Зависимость результата измерения угла наклона от высоты

полнено моделирование системы измерения параметров при следующих условиях: источник излучения Cz-137 с энергией 660 кэВ, расстояние $l = 2i$, начальная высота $h = 8i$, угол коллимации 60 град.

На рис. 3 показан пример измерения с использованием рекуррентной нейронной сети высоты спускаемого аппарата, движущегося равномерно, по данным моделирования фотонного преобразователя. Моделирование выполнялось при разных значениях угла наклона спускаемого аппарата $\gamma_1 = 0, -\pi/64, -\pi/32, -3\pi/64, -\pi/16$ рад. Средняя величина относительной погрешности в данном эксперименте составила 1 %.

На рис. 4 приведен график зависимости результата измерения угла наклона оси спускаемого аппарата от высоты в случае, если угол наклона не изменяется в процессе спуска. Моделировались значения угла наклона $[0, -\pi/64, -\pi/32, -3\pi/64, -\pi/16]$ рад. Относительная погрешность измерения угла наклона в данном эксперименте (за исключением

случая $\gamma_1 = 0$) составила 4 %.

Разработанная компьютерная модель системы измерения параметров движения спускаемого аппарата на поверхность Луны предназначена для использования на начальных этапах разработки системы, поскольку выполнение полного натурального эксперимента невозможно.

Предложено использовать нелинейную модель пространства состояний, которая лучше характеризует неравномерное движение вблизи лунной поверхности при отсутствии атмосферы.

Нелинейная модель пространства состояний реализуется рекуррентной нейронной сетью с задержками на входе.

Компьютерный эксперимент показал, что при выбранных конструктивных параметрах системы нейросетевой алгоритм на этапе обучения позволяет получить среднюю относительную погрешность измерения высоты 1 %, угла наклона — 4 %, на этапе обобщения 1,5 и 6 %.

СПИСОК ЛИТЕРАТУРЫ

1. Луна — шаг к технологиям освоения Солнечной системы. Под ред. В.П. Легостаева и В.А. Лопоты. М.: РКК «Энергия», 2011. 684 с.
2. Поляков В.М., Покровский В.П., Сомс Л.Н. Лазерный передающий модуль с переключаемой диаграммой направленности для дальномера космического аппарата «Фобос-грунт». 2011.

3. Зубов Н.Е., Савчук Д.В., Старовойтов Е.И. Анализ возможностей, оптимизации массы и энергопотребления лазерного высотомера для управления спуском с окололунной орбиты // Космическая техника и технологии. 2014. № 1 (4). С. 67–74.
4. ЗАО «Скат-Р» [электронный ресурс]/ URL:

<http://www.skat-r.ru> (дата обращения: 15.05.2015).

5. **Егоров В.В.** Доплеровские радары посадки космических аппаратов на Луну и планеты Солнечной системы // Современные проблемы дистанционного зондирования Земли из космоса. 2012. Т. 9. № 2. С. 145–151

6. **Ковалев М.А. и др.** Радиотехнические навигационные средства самолетовождения и посадки: учеб. пособие. В 2 ч. Ч. 1. Самара: Изд-во Самар. гос. аэрокосм. ун-та, 2011. 76 с.

7. УПКБ «Деталь» [электронный ресурс]/ URL:

<http://www.upkb.ru/> (дата обращения: 15.05.2015).

8. ОАО «Корпорация «Тактическое ракетное вооружение» [электронный ресурс]/ URL: <http://ktrv.ru/> (дата обращения: 15.05.2015).

9. **Юревич Е.И.** Основы проектирования техники: учеб. пособие. СПб.: Изд-во СПбГПУ, 2012. 148 с.

10. **Васильев А.Н., Тархов Д.А.** Построение приближённых нейросетевых моделей по разнородным данным // Математическое моделирование. 2007. № 19:12. С. 43–51.

REFERENCES

1. **Legostayev V.P., Lopota V.A.** (eds.). *Luna – shag k tekhnologiyam osvoyeniya Solnechnoy sistemy* [The Moon – a Step Towards the Development of Technologies of the Solar System], Moscow: RKK «Energiya» Publ., 2011, 684 p. (rus)

2. **Polyakov V.M., Pokrovskiy V.P., Soms L.N.** *Lazernyy peredayushchiy modul s pereklyuchayemoy diagrammoy napravlenosti dlya dalnomera kosmicheskogo apparata «Fobos-grunt»* [The Laser Transmitter Module with Switchable Radiation Pattern for EDM Spacecraft “Phobos-Grunt”]. 2011. (rus)

3. **Zubov N.Ye., Savchuk D.V., Starovoytov Ye.I.** Analiz vozmozhnostey, optimizatsii massy i energopotrebleniya lazernogo vysotomera dlya upravleniya spuskom s okololunnoy orbity [Analysis of Possibilities, Optimization of Mass and of Power Consumption for a Laser Altimeter Controlling the Descent of a Spacecraft from the Lunar Orbit]. *Kosmicheskaya tekhnika i tekhnologii* [Space technic and technology], 2014, No. 1 (4), Pp. 67–74. (rus)

4. ЗАО «Skat-R». Available: <http://www.skat-r.ru> (Accessed: 15.05.2015). (rus)

5. **Yegorov V.V.** Doplerovskiye radary posadki kosmicheskikh apparatov na Lunu i planety Solnechnoy sistemy [Doppler Radars for the

Spacecrafts Landing on Moon and Planets of Solar System]. *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa* [Current Problems in Remote Sensing of the Earth from Space], 2012, Vol. 9, No. 2, Pp.145–151. (rus)

6. **Kovalev M.A. et al.** *Radiotekhnicheskiye navigatsionnyye sredstva samoletovozhdeniya i posadki* [Radio-technical Navigation Aids and Aircraft Navigation Landing], Part 1. Samara: Samar. gos. aerokosm. un-t Publ., 2011, 76 p. (rus)

7. УПКБ «Деталь». Available: <http://www.upkb.ru> (Accessed: 15.05.2015). (rus)

8. ОАО «Корпорация «Тактическое ракетное вооружение». Available: <http://ktrv.ru/> (Accessed: 15.05.2015). (rus)

9. **Yurevich Ye.I.** *Osnovy proyektirovaniya tekhniki* [Basics of designing engineering], St. Petersburg: SPbGPU Publ., 2012, 148 p. (rus)

10. **Vasilyev A.N., Tarkhov D.A.** Postroyeniye priblizhennykh neyrosetevykh modeley po raznorodnym dannym [Construction of approximate neural network models according to heterogeneous data]. *Matematicheskoe Modelirovaniye* [Mathematical Models and Computer Simulations], 2007, Vol. 19:12, Pp. 43–51. (rus)

МАЛЫХИНА Галина Федоровна – профессор кафедры информационных измерительных технологий Санкт-Петербургского политехнического университета Петра Великого, доктор технических наук.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: g_f_malychina@mail.ru

MALYKHINA Galina F. Peter the Great St. Petersburg Polytechnic University.

195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.

E-mail: g_f_malychina@mail.ru

КИСЛИЦЫНА Ирина Александровна – аспирант кафедры информационных измерительных технологий Санкт-Петербургского политехнического университета Петра Великого.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: irina_kislitsyna@mail.ru

KISLITSINA Irina A. Peter the Great St. Petersburg Polytechnic University.

195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.

E-mail: irina_kislitsyna@mail.ru

DOI: 10.5862/JCSTCS.229.7

УДК 004

К.В. Соколов, Д.А. Тимофеев, А.В. Самочадин

ИЗВЛЕЧЕНИЕ ОПИСАНИЙ БИЗНЕС-ПРОЦЕССОВ ИЗ ТЕКСТОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ

K.V. Sokolov, D.A. Timofeev, A.V. Samochadin

EXTRACTION OF BUSINESS PROCESS MODELS FROM TEXTS

Для автоматизации бизнес-процессов необходимо наличие их формальных моделей. Создание таких моделей – трудоемкий процесс, требующий привлечения квалифицированных экспертов. В то же время для многих процессов уже существуют описания на естественном языке в виде регламентов, инструкций, положений и других официальных документов, а также в виде менее формализованных описаний, составляемых участниками процессов. Рассмотрена методика извлечения описаний бизнес-процессов из набора текстов и используемые для этого алгоритмы. Отличительной особенностью предлагаемого подхода является применение методов глубинного анализа процессов (process mining).

БИЗНЕС-ПРОЦЕССЫ; ИЗВЛЕЧЕНИЕ ПРОЦЕССОВ; ГЛУБИННЫЙ АНАЛИЗ ПРОЦЕССОВ; СЕМАНТИЧЕСКИЙ АНАЛИЗ; ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА.

Formal models of business processes are necessary for process analysis and for developing process-oriented services. Process modeling is a time-consuming task which should be performed by an expert. A large part of the expert's work consists of gathering information about the process. In most cases, process descriptions already exist in textual form. The automated method of extracting process models from texts is required to put this knowledge into use. In this paper we develop a process extraction method for texts like records, reports, or user stories. The main feature of the proposed method is the application of process mining techniques to text analysis.

BUSINESS PROCESS; PROCESS EXTRACTION; PROCESS MINING; SEMANTIC ANALYSIS; NATURAL LANGUAGE PROCESSING.

Анализ и автоматизация бизнес-процессов требуют использования их формальных моделей. Разработка формальных описаний процессов – это трудоемкая задача, требующая привлечения экспертов-аналитиков. Как следствие, актуальной является задача автоматизации моделирования процессов.

В ходе работы аналитик использует доступные источники информации об исследуемом процессе. Во многих случаях

такими источниками служат тексты: спецификации, регламенты, инструкции, методические указания. Кроме официальных документов могут использоваться тексты, составленные самими участниками процесса, такие как отчеты или отзывы, описывающие конкретный опыт участника. В некоторых случаях (например, при телефонном обращении в службу технической поддержки или в органы государственной власти) могут записываться сеансы взаимодействия

с клиентами. Средства автоматизированного анализа таких источников позволили бы не только сократить временные затраты на построение формального описания процесса, но и повысить полноту анализа за счет увеличения количества информации о процессе.

Одной из областей, в которых автоматизированное извлечение описаний бизнес-процессов из текстов представляется особенно полезным, является разработка процессно-ориентированных мобильных сервисов. Пример такого рода сервисов – мобильные помощники, подсказывающие пользователю последовательность шагов, которые необходимо выполнить для достижения цели.

В университетской среде такое приложение может способствовать адаптации недавно поступивших студентов, предоставляя им персонализированную информацию о порядке оформления в качестве студента и доступе к ресурсам университета (общежития, библиотеки, спортивные секции), а также помогать выполнению студентами и сотрудниками стандартных процедур (оформление академического отпуска, восстановление, оформление командировок, участие в конкурсах на получение грантов). Каждому такому процессу соответствует общая модель, единая для всех процессов соответствующего класса, и частная модель, учитывающая особенности данного процесса: например, запись в спортивные секции выполняется по общим правилам, но для каждой конкретной секции детали процесса записи могут отличаться.

Поскольку частные особенности конкретных процессов не всегда описаны в официальных документах и могут со временем меняться, представляется полезным уточнять модели процессов на основе обратной связи от самих участников процессов. Не будучи экспертами в моделировании бизнес-процессов, пользователи могли бы уточнять модели процессов путем отправки текстов, описывающих их личный опыт реализации соответствующего процесса. Возможность анализа таких частичных описаний процессов и объединения извлеченной из них информации с существующими мо-

делями позволила бы лучше адаптировать систему к изменениям процессов.

Известные подходы к извлечению описаний процессов

Выделяются два класса работ, посвященных автоматическому выводу моделей процессов. Первый, наиболее обширный класс, посвящен задаче извлечения процессов из протоколов работы информационных систем, известной как задача глубинного анализа процессов (process mining) [1].

Протокол представляет собой последовательность событий, каждому из которых сопоставлен идентификатор реализации процесса, одинаковый для всех событий, которые имели место при выполнении процесса в конкретном случае. Например, при анализе процесса обслуживания клиентов интернет-магазина в качестве такого идентификатора можно использовать номер заказа. Часто каждая запись протокола также содержит идентификатор пользователя, выполнившего операцию, и отметку о времени события.

Протоколы описывают процесс только частично, поскольку некоторые необязательные действия могут не иметь места в каждом конкретном случае. Часто протоколы также содержат ошибки, которые могут быть связаны, например, с некорректным выполнением инструкции оператором или с отсутствием части данных.

Цель алгоритмов глубинного анализа – построение формальной модели процесса, наилучшим образом соответствующей данному протоколу. Процессы при этом чаще всего представляются в виде графов переходов или сетей Петри. Область глубинного анализа процессов активно развивается, однако предметом исследований являются в первую очередь методы и алгоритмы анализа протоколов; задача извлечения формальных описаний процессов из данных другой природы, насколько нам известно, не рассматривается.

Второй класс работ посвящен извлечению описаний процессов непосредственно из текстов. Основным подходом, которого придерживается большинство авторов, является выделение в тексте элементов,

составляющих бизнес-процесс, и формирование на их основе описания процесса в соответствии с базовой концептуальной моделью.

В качестве базовой модели часто используется модель CREWS [2]. В ней составляющими процесса являются объекты, которые могут выступать в качестве действующих лиц и ресурсов, и действия, элементарные или составные. Аналогом этих понятий в лингвистическом контексте выступает аргументно-предикатная структура предложения. Выделение сущностей и отношений – стандартная задача обработки естественного языка. Основные отличия между известными подходами к извлечению описаний процессов состоят в используемых авторами методах анализа текста.

В системе R-BPD [3] для извлечения составляющих процесса используется поиск на основе синтаксических шаблонов, определение частей речи и статистическое выделение компонентов предложения.

В работе [4] для анализа сценариев использования реализован более сложный алгоритм, включающий поверхностный синтаксический анализ, идентификацию словарных слов и словосочетаний, разрешение анафоры и анализ контекста.

Похожий метод использован в работах [5, 6], посвященных автоматизированному анализу пользовательских историй.

В работе [7] для анализа текста применен статистический парсер [8], а в качестве источника информации о семантических отношениях (в частности, для учета синонимии) использованы базы данных WordNet [9] и FrameNet [10].

С точки зрения используемых методов выделяется работа [11], авторы которой используют для описания моделей рекурсивные шаблоны, представленные в виде структур признаков. Аналогичные структуры признаков используются на синтаксическом уровне при анализе текста. Поиск соответствия между шаблонами и синтаксическими конструкциями реализуется путем унификации структур признаков.

В работе [12] рассматривается случай линейных процессов, каждое действие в которых описывается глаголом. Модель

процесса строится путем анализа отношения предшествования на множестве пар, состоящих из глагола и следующего за ним существительного (предполагая, что в большинстве случаев такое существительное является прямым дополнением), которые извлекаются из набора документов. Особенность этой работы – метрика соответствия документа линейной структуре процесса.

К извлечению описаний процессов близка задача оценки степени отклонения описания процесса на естественном языке от формальной модели, рассматриваемая в работе [13]. Близость описаний оценивается путем определения соответствия между фрагментами текста, которым сопоставляются множества наиболее информативных слов, и компонентами формальной модели, для которых аналогичные множества слов извлекаются из онтологии.

Анализ и постановка задачи

Рассмотрим, какие сложности возникают в указанных подходах при попытке использовать тексты на естественном языке.

С точки зрения структуры описания процесса можно выделить два класса документов. К документам первого класса относятся инструкции, спецификации, регламенты. Их общая черта – рассмотрение различных аспектов процесса и вариантов его выполнения. Второй класс документов включает описания отдельных реализаций процесса. Примерами таких документов являются протоколы, отчеты, некоторые виды отзывов.

Извлечение описаний процессов из документов типа инструкций затруднительно. Анализ документов сложной структуры требует обращения к методам лингвистики текста. Помимо линейной последовательности текст характеризуется существованием структурных единиц особого типа (элементарная дискурсивная единица, сверхфразовое единство и пр.) и ряда взаимосвязанных нелинейных структур, многие из которых представлены в тексте неявно (ср. понятия когезии и когерентности, топика, актуального членения, макроструктуры и макропропозиции, риторической структуры и

др.), а также нетривиальным соотношением локальных и глобальных языковых явлений (пресуппозиции и имплицатуры, контекст, речевые акты). Рассматриваемый подход к построению модели процесса потребует выявления основных единиц текста и экспликации подобного рода структур во всей их взаимосвязи в качестве модели текста, на основе которого должна восстанавливаться модель описываемого в этом тексте процесса. Известные в настоящий момент попытки построения формальных моделей на основе текстов данного типа ориентированы на идентификацию фрагментов заранее определенной модели с использованием извлекаемого из текстов фактического материала (например, [14]). Полностью автоматическая реконструкция процесса по изолированному подробному текстовому описанию нетривиальной структуры в настоящее время не представляется возможной (но ср. работы по автоматическому реферированию [15], машинному переводу и автоматическому пониманию текстов [16])

Тексты второго класса (отчеты и протоколы) значительно проще по структуре, хотя также могут содержать нетривиальные конструкции как на уровне предложения, так и на уровне всего текста в целом. Здесь основным содержанием текста является последовательность действий, выполненных в конкретном случае. Это свойство сближает такие тексты с протоколами работы информационных систем.

Формализация семантики такого рода текстов может опираться на иную концепцию значения, отличающуюся от традиционной для статической формальной семантики концепции значения как условий истинности. В данном случае интерпретация текста может опираться на последовательность сообщений, модифицирующих информированность адресата (т. н. принцип «потенциала изменения контекста»), и носит выраженный динамический характер. В рамках динамической семантики известны такие подходы, как теория репрезентации дискурса и динамическая логика предикатов. Данная концепция значения позволяет предложить решения ряда упоминавшихся выше проблем, в частности, разрешения

анафоры как одного из механизмов обеспечения связности текста [17].

Если основная информация в тексте будет соответствующим образом формализована, для извлечения модели процесса можно использовать алгоритмы глубинного анализа процессов. В данной статье мы рассматриваем реализацию этого подхода. Мы начинаем с рассмотрения частного случая: документов, описывающих конкретную последовательность действий, которая должна быть выполнена — или действительно была выполнена автором — для достижения цели.

На первом этапе должна быть решена задача преобразования текста, описывающего последовательность действий, в структурированный протокол, к которому можно применять алгоритмы извлечения процессов. В этой статье мы рассматриваем алгоритм анализа текстов, в которых описание последовательности действий усложняется языковыми конструкциями «если — то», «и», «или», выражающими их зависимость, совместность и альтернативность.

В настоящее время мы решаем задачу извлечения описаний процессов только для английского языка, что связано в первую очередь с доступностью лингвистических ресурсов. Тем не менее мы строим алгоритмы таким образом, чтобы разработанный метод мог использоваться для анализа текстов на русском и других языках.

Предлагаемое решение

Мы предлагаем следующую последовательность шагов для извлечения формального описания процесса из текста.

1. Формирование семантического представления для каждого предложения с помощью семантического парсера.

2. Идентификация сущностей в полученных представлениях.

3. Генерация линейной последовательности событий в формате протокола.

4. Формирование соответствующей протоколу модели процесса в виде сети Петри.

Использование семантического формализма позволяет отвлечься от особенностей конкретного языка. В качестве такого формализма мы используем абстрактное



представление значения (Abstract Meaning Representation, AMR) [18]. AMR представляет собой ориентированный граф с выделенной вершиной. Каждому узлу графа соответствует сущность, а ребрам сопоставляются атрибуты, описывающие отношения между сущностями. Набор атрибутов в AMR представляет собой расширенный набор семантических ролей PropBank [19]. В частности, в AMR представимы интересующие нас конструкции условия, объединения и альтернативы, которым сопоставлены атрибуты «condition», «and» и «or» соответственно. Семантический парсер — это программа, которая принимает на вход последовательность предложений и для каждого из них возвращает граф AMR, представляемый обычно в виде списочной структуры формата ((метка / узел) :атрибут подграф ... :атрибут подграф), где в качестве подграфа может выступать аналогичная списочная структура или метка узла. Для AMR существует несколько семантических парсеров; мы используем парсер, описанный в работе [20].

Генерируемый на основе текста протокол представляется в виде списка пар вида (i, P_i) , где i — номер реализации, а $P_i = [A_1; \dots; A_p]$ — упорядоченный список действий, имеющих место в данной реализации процесса. Мы предполагаем, что действия выполняются в том порядке, в котором следуют предложения в тексте, а порядок совместных действий соответствует порядку их упоминания в предложении. Для упрощения описания алгоритма положим, что если s и t — семантические представления действий, то конструкция $(if\ st)$ описывает зависимость («если выполнено действие s , то нужно выполнить действие

t »), конструкция $(or\ st)$ описывает альтернативность («выполните действие s или действие t »), конструкция $(and\ st)$ описывает совместность («выполните действие s и действие t »).

Алгоритм формирования протокола

1. Выделить в графе подграфы, соответствующие элементарным, зависимым, совместным и альтернативным действиям.

2. Нормализовать представления совместно и альтернативно реализуемых действий. Подграфы, соответствующие конструкциям «выполнить действие s с объектами a и b », преобразуются в пары подграфов «выполнить действие s с объектом a » и «выполнить действие s с объектом b ». Преобразование альтернативных конструкций производится аналогично.

3. Для каждого элементарного действия сгенерировать текстовое описание, которое станет элементом протокола.

4. Сформировать протокол путем последовательного добавления в него действий в порядке их следования в тексте и с учетом их структуры.

Добавление действий в протокол можно описать следующим псевдокодом (см. листинг 1). Здесь квадратные скобки обозначают список, знак «+» обозначает конкатенацию списков, конструкция $[case + x \mid case \leftarrow Protocol]$ обозначает создание нового протокола, который строится из существующего протокола $Protocol$ путем добавления в конец каждой входящей в протокол реализации $case$ последовательности действий x . Конструкция $match\ s\ with$ обозначает сопоставление с образцом: она выбирает тот вариант действия, который соответствует структуре анализируемого объекта s .

```
def AddAction(Action, Protocol) =
  match Action with
  | (and s1 s2) => [case + [s1; s2] | case <- Protocol]
  | (or s1 s2)  => [case + [s1] | case <- Protocol] +
                  [case + [s2] | case <- Protocol]
  | (if s1 s2)  => [case + [s1; s2] | case <- Protocol] +
                  Protocol
  | s           => [case + [s] | case <- Protocol]
end
```

Листинг 1. Алгоритм добавления действий в протокол

Как видно из приведенного псевдокода, альтернативные и зависимые конструкции увеличивают количество реализаций в протоколе: при обработке альтернативной конструкции мы берем все построенные к этому моменту реализации процесса и для каждой из них строим все возможные варианты выполнения действия. Аналогично для зависимых действий рассматриваются

два варианта развития событий: выполнение условия и его невыполнение.

Текстовые описания элементарных действий строятся на основе соответствующих им подграфов с помощью фиксированного набора правил. Генерация полноценных описаний действий является предметом будущего исследования.

Для вывода формального описания про-

```
# ::snt Check your department's deadlines.
(x1 / check-01
  :ARG1 (x5 / deadline
        :poss (x3 / department)))

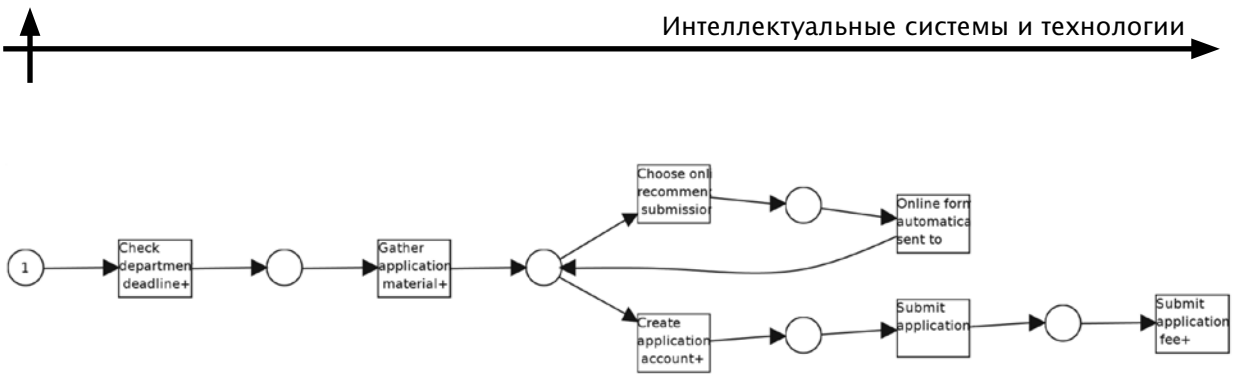
# ::snt Gather necessary application materials.
(x1 / gather-01
  :ARG1 (x4 / material
        :mod (x3 / apply-01)))

# ::snt If you choose online recommendation submission, the online form will
be automatically sent to recommenders.
(x14 / send-01
  :condition (x3 / choose-01
    :ARG1 (x6 / submission
      :mod (x5 / recommend-01
        :mod (x4 / online))))
  :ARG1 (x10 / form
    :mod (x9 / online))
  :manner (x13 / automatically)
  :ARG2 (x16 / recommender))

# ::snt Create an application account and submit the application, and the
application fee.
(x5 / and
  :op1 (x1 / create-01
    :ARG1 (x4 / account
      :mod (x3 / apply-01)))
  :op2 (x6 / submit-01
    :ARG1 (x10 / and
      :op1 (x8 / apply-01)
      :op2 (x13 / fee
        :ARG1 (x12 / apply-01))))))

# ::snt Submit official transcripts from each institution you have attended.
(x1 / submit-01
  :ARG1 (x3 / transcript
    :mod (x2 / official)
    :source (x6 / institution
      :mod (x5 / each)
      :ARG1-of (x9 / attend-01))))
```

Листинг 2. Соответствующие анализируемым предложениям структуры AMR



Выведенная модель процесса

цесса мы используем систему глубинного анализа процессов ProM (<http://promtools.org>), в частности, плагин Mine for a Petri Net using ILP. Анализ применимости и эффективности различных алгоритмов извлечения формальных моделей из протоколов, построенных на основе текста, также остается предметом дальнейшей работы.

Рассмотрим в качестве примера фрагмент текста, описывающий один из этапов подачи документов в университет.

Check your department's deadlines. Gather necessary application materials. If you choose online recommendation submission, the online form will be automatically sent to recommenders. Create an application account and submit the application, and the application fee.

В ходе семантического анализа этих предложений формируется набор структур AMR, показанный в листинге 2.

Применяя к набору графов алгоритм формирования протокола, получаем две реализации процесса: в одной из них действие «choose online recommendation submission» не выполняется, в другом — выполняется, и вслед за этим выполняется обязатель-

ное в этом случае действие «online form automatically sent to recommenders». Итоговый протокол показан в листинге 3.

Подав этот протокол на вход системы ProM, получаем модель процесса в виде сети Петри, приведенной на рисунке.

Открытые вопросы и дальнейшие исследования

Создание метода извлечения формальных описаний процессов из текста даже в ограниченной постановке требует решения ряда проблем. Перечислим наиболее важные из них.

Даже если ограничиться анализом одного документа, одни и те же сущности нередко могут обозначаться в тексте несколькими разными способами (синонимичные именные группы, местоимения в анафорическом употреблении). Задача становится еще более трудной при работе с несколькими текстами, написанными разными авторами. Необходим систематический способ поиска похожих описаний и проверки гипотезы об их соответствии одной и той же сущности. Необходим также метод со-

CaseId	Action
1	Check department's deadline
1	Gather application material
1	Choose online recommendation submission
1	Online form is automatically sent to recommender
1	Create application account
1	Submit application
1	Submit application fee
2	Check department's deadline
2	Gather application material
2	Create application account
2	Submit application
2	Submit application fee

Листинг 3. Извлеченная из текста последовательность действий

гласования частичных описаний процесса, включающих только подмножества путей, по которым может идти выполнение, и этапов, из которых состоит полный процесс.

Разработанный алгоритм опирается на существенное ограничение: мы полагаем, что все действия выполняются в порядке их описания в тексте. Обобщение полученных нами результатов на более широкий класс текстов требует более точного анализа временной последовательности событий. Еще более важным оказывается упорядочение событий при анализе нескольких описаний одного процесса.

В данной статье мы ограничились рассмотрением простых процессов, состоящих из элементарных действий или событий. Интересен вопрос, насколько предлагаемый нами подход может быть обобщен на случай иерархических процессов, в качестве действий в которых могут выступать вложенные процессы, а также на случай, когда текст содержит описание нескольких взаимосвязанных или одновременно выполняющихся процессов.

Наконец, одной из наиболее важных проблем является формирование метрик, позволяющих сравнивать модели процессов и оценивать качество разрабатываемых алгоритмов. В работе [21] приведен ряд метрик, используемых при оценке результатов работы алгоритмов глубинного анализа процессов, но эти метрики оценивают соответствие процессов протоколам, полученным при реальном выполнении процесса. В нашем случае оценка качества рабо-

ты алгоритма может производиться путем сравнения выведенного описания процесса с эталоном, построенным экспертом-аналитиком. При этом необходимо учитывать, что синтаксического сравнения моделей недостаточно: один и тот же процесс может быть описан несколькими различными, но эквивалентными способами. Более точными представляются метрики, основанные на сравнении множеств протоколов, допускаемых каждым описанием. Если рассматривать модель процесса как грамматику, процессы эквивалентны, если они порождают один и тот же язык. Исследование методов сравнения моделей процессов представляет собой одну из наиболее приоритетных задач дальнейшей работы.

В настоящей статье предложен метод построения формальных описаний процессов путем анализа текстов на английском языке, описывающих выполняемые в рамках каждого процесса действия. Отличительной особенностью представленного подхода является применение методов глубинного анализа процессов (process mining). В статье также рассмотрены некоторые открытые вопросы, решение которых позволит обобщить полученные результаты на более широкий класс описаний процессов.

Работа выполнена в ходе реализации комплексного проекта в рамках Постановления Правительства РФ от 09.04.2010 №218 при финансовой поддержке Минобрнауки России. Договор № 02.G25.31.0024 от 12.02.2013.

СПИСОК ЛИТЕРАТУРЫ

1. **van der Aalst W.M.P.** Process Mining: Discovery, Conformance and Enhancement of Business Processes. Berlin: Springer-Verlag, 2011.
2. **Achour C.B.** Guiding Scenario Authoring // European-Japanese Conf. on Information Modelling and Knowledge Bases. 1998.
3. **Ghose A., Koliadis G., Chueng A.** Process discovery from model and text artefacts // IEEE Congress on Services. 2007. Pp. 167–174.
4. **Sinha A., Paradkar A., Kumanan P., Boguraev B.** An Analysis Engine for Dependable Elicitation on Natural Language Use Case Description and its Application to Industrial Use Cases // IBM Research Report RC24712 (W0812-106). 2008.
5. **Gonçalves J.C.A.R., Santoro F.M., Baião F.A.** Business Process Mining from Group Stories // Proc. of the 13th Internat. Conf. on Computer Supported Cooperative Work in Design. 2009.
6. **Gonçalves J.C.A.R., Santoro F.M., Baião F.A.** Let Me Tell You a Story – On How to Build. Process Models // J. of Universal Computer Science. 2011. Vol. 17. Pp. 276–295.
7. **Friedrich F., Mendling J., Puhmann, F.** Process model generation from natural language text // Advanced Information Systems Engineering. 2011. Pp. 482–496.
8. **Manning D.** Natural Language Parsing. Advances in Neural Information Processing Systems



// Proc. of the 2002 Conf. 2003. Vol. 15.

9. **Miller G.A.** WordNet: a lexical database for English // *Communications of the ACM*. 1995. Vol. 38. Pp. 39–41.

10. **Baker C.F., Fillmore C.J., Lowe J.B.** The Berkeley FrameNet Project // Proc. of the 17th Internat. Conf. on Computational Linguistics. Association for Computational Linguistics. 1998. Vol. 1. Pp. 86–90.

11. **Ackermann L., Volz B.** Model [NL] generation: natural language model extraction // Proc. of the 2013 ACM Workshop on Domain-Specific Modeling. 2013. Pp. 45–50.

12. **Chen Y., Ding Zh., Sun H.** PEWP: Process Extraction Based on Word Position in Documents // Proc. of the 9th Internat. Conf. on Digital Information Management (ICDIM). 2014. Pp. 135–140.

13. **van der Aa H., Leopold H., Reijers H.A.** Detecting Inconsistencies Between Process Models and Textual Descriptions // *Business Process Management*. Springer, 2015. Pp. 90–105.

14. **Novichkova S., Egorov S., Daraselia N.** MedScan, a natural language processing engine for MEDLINE abstracts // *Bioinformatics*. 2003. Vol. 19. No. 13. Pp. 1699–1706.

15. **Marcu D.** The rhetorical parsing of

unrestricted texts: A surface-based approach // *Computational Linguistics*. 2000. Vol. 26. No. 3. Pp. 395–448.

16. **Леонтьева Н.Н.** Автоматическое понимание текстов. Системы, модели, ресурсы. М.: Академия, 2006.

17. **Gerasimov N., Pyshkin E.** Using Dynamic Predicate Logic for Pronominal Anaphora Resolution in Russian Texts // Proc. of the Internat. Workshop on Applications in Information Technology. Aizu-Wakamatsu, Japan, 2015. Pp. 51–54.

18. **Banarescu L. et al.** Abstract Meaning Representation for Sembanking // Proc. of Linguistic Annotation Workshop. 2013.

19. **Palmer M., Kingsbury P., Gildea D.** The Proposition Bank: An Annotated Corpus of Semantic Roles // *Computational Linguistics*. 2005. Vol. 31. No. 1. Pp. 71–106.

20. **Wang C., Xue N., Pradhan S., Pradhan S.** A Transition-based Algorithm for AMR Parsing // *Human Language Technologies: Annual Conf. of the North American Chapter of the ACL*. 2015. Pp. 366–375.

21. **Rozinat A., Veloso M., van der Aalst W.M.P.** Evaluating the Quality of Discovered Process Models // Proc. of the ECML-PKDD Workshop on Induction of Process Models. 2008. Pp. 45–52.

REFERENCES

1. **van der Aalst W.M.P.** *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Berlin: Springer-Verlag, 2011.

2. **Achour C.B.** Guiding Scenario Authoring. *European-Japanese Conference on Information Modelling and Knowledge Bases*, 1998.

3. **Ghose A., Koliadis G., Chueng A.** Process discovery from model and text artefacts. *IEEE Congress on Services*, 2007, Pp. 167–174.

4. **Sinha A., Paradkar A., Kumanan P., Boguraev B.** An Analysis Engine for Dependable Elicitation on Natural Language Use Case Description and its Application to Industrial Use Cases. *IBM Research Report RC24712 (W0812-106)*, 2008.

5. **Gonçalves J.C.A.R., Santoro F.M., Baião F.A.** Business Process Mining from Group Stories. *Proceedings of the 13th International Conference on Computer Supported Cooperative Work in Design*, 2009.

6. **Gonçalves J.C.A.R., Santoro F.M., Baião F.A.** Let Me Tell You a Story – On How to Build Process Models. *Journal of Universal Computer Science*, 2011, Vol. 17, Pp. 276–295.

7. **Friedrich F., Mendling J., Puhmann F.** Process model generation from natural language text. *Advanced Information Systems Engineering*, 2011, Pp. 482–496.

8. **Manning D.** Natural Language Parsing. *Advances in Neural Information Processing Systems, Proceedings of the 2002 Conference*, 2003, Vol. 15.

9. **Miller G.A.** WordNet: a lexical database for English. *Communications of the ACM*, 1995, Vol. 38, Pp. 39–41.

10. **Baker C.F., Fillmore C.J., Lowe J.B.** The Berkeley FrameNet project. *Proceedings of the 17th International Conference on Computational linguistics. Association for Computational Linguistics*, 1998, Vol. 1, Pp. 86–90.

11. **Ackermann, L., Volz B.** Model [NL] generation: natural language model extraction. *Proceedings of the 2013 ACM workshop on domain-specific modeling*, 2013, Pp. 45–50.

12. **Chen Y., Ding Zh., Sun H.** PEWP: Process Extraction Based on Word Position in Documents. *Proceedings of the 9th International Conference on Digital Information Management*, 2014, Pp. 135–140.

13. **van der Aa H., Leopold H., Reijers H.A.** Detecting Inconsistencies Between Process Models and Textual Descriptions. *Business Process Management*, Springer, 2015, Pp. 90–105.

14. **Novichkova S., Egorov S., Daraselia N.** MedScan, a natural language processing engine for MEDLINE abstracts. *Bioinformatics*, 2003, Vol. 19,

No. 13, Pp. 1699–1706.

15. **Marcu D.** The rhetorical parsing of unrestricted texts: A surface-based approach. *Computational Linguistics*, 2000, Vol. 26, No. 3, Pp. 395–448.

16. **Leontyeva N.N.** *Avtomaticheskoye ponimaniye tekstov. Sistemy, modeli, resursy* [Automatic understanding of texts. Systems, models, resources]. Moscow: Akademiya Publ., 2006. (rus)

17. **Gerasimov N., Pyshkin E.** Using Dynamic Predicate Logic for Pronominal Anaphora Resolution in Russian Texts. *Proceedings of the International Workshop on Applications in Information Technology*, Aizu-Wakamatsu, Japan, 2015, Pp. 51–54.

18. **Banarescu L. et al.** Abstract Meaning

Representation for Sembanking. *Proceedings of Linguistic Annotation Workshop*, 2013.

19. **Palmer M., Kingsbury P., Gildea D.** The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 2005, Vol. 31, No. 1, Pp. 71–106.

20. **Wang C., Xue N., Pradhan S., Pradhan S.** A Transition-based Algorithm for AMR Parsing. *Human Language Technologies: Annual Conference of the North American Chapter of the ACL*, 2015, Pp. 366–375.

21. **Rozinat A., Veloso M., van der Aalst W.M.P.** Evaluating the Quality of Discovered Process Models. *Proceedings of the ECML-PKDD Workshop on Induction of Process Models*, 2008, Pp. 45–52.

СОКОЛОВ Константин Владимирович — аспирант кафедры математической лингвистики Санкт-Петербургского государственного университета.

199034, Россия, Санкт-Петербург, Университетская наб., д. 7-9.

E-mail: vtqveant@gmail.com

SOKOLOV Konstantin V. *St. Petersburg State University.*

199034, Universitetskaya Emb. 7-9, St. Petersburg, Russia.

E-mail: vtqveant@gmail.com

ТИМОФЕЕВ Дмитрий Андреевич — старший преподаватель кафедры распределенных вычислений и компьютерных сетей Института компьютерных наук и технологий Санкт-Петербургского политехнического университета Петра Великого.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: dtim@dcn.icc.spbstu.ru

TIMOFEEV Dimitri A. *Peter the Great St. Petersburg Polytechnic University.*

195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.

E-mail: dtim@dcn.icc.spbstu.ru

САМОЧАДИН Александр Викторович — профессор кафедры распределенных вычислений и компьютерных сетей Института компьютерных наук и технологий Санкт-Петербургского политехнического университета Петра Великого, кандидат технических наук.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: samochadin@gmail.com

SAMOCHADIN Alexander V. *Peter the Great St. Petersburg Polytechnic University.*

195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.

E-mail: samochadin@gmail.com



DOI: 10.5862/JCSTCS.229.8

УДК 025.4, 004

*В.Н. Корелин, И.С. Блеканов, С.Л. Сергеев***ПРИМЕНЕНИЕ МОДИФИЦИРОВАННОГО АЛГОРИТМА LSH
ДЛЯ КЛАСТЕРИЗАЦИИ ВНЕШНЕГО ОКРУЖЕНИЯ ВЕБ-ПРОСТРАНСТВА
УНИВЕРСИТЕТОВ****V.N. Korelin, I.S. Blekanov, S.L. Sergeev***CLUSTERING OF THE EXTERNAL WEB ENVIRONMENT OF UNIVERSITIES
USING A MODIFIED LSH ALGORITHM**

Проведен кластерный анализ внешних ресурсов сайтов крупных университетов. В качестве исследуемых объектов выбраны сайты университетов России, США и Великобритании, занимающие в своих регионах ведущие позиции в вебметрическом рейтинге. Цель работы – в выявлении для каждого сайта университета групп внешних веб-ресурсов с одинаковым родом деятельности. Проведен анализ найденных групп: определена степень влияния количества и размеров этих групп на вебметрический рейтинг сайтов университетов. Разработан алгоритм кластеризации, основанный на вероятностном методе понижения размерности многомерных данных (Locality-Sensitive Hashing – LSH). Поставлен эксперимент, в котором на тестовых данных показано, что алгоритм позволяет с высокой скоростью и допустимой точностью проводить кластеризацию большого объема данных. Приведены основные результаты исследования.

ВЕБОМЕТРИКА; ВЕБ-САЙТЫ УНИВЕРСИТЕТОВ; КЛАСТЕРНЫЙ АНАЛИЗ; LOCALITY-SENSITIVE HASHING; MIN HASHING; КЛАСТЕРИЗАЦИЯ ВНЕШНИХ ВЕБ-РЕСУРСОВ; АНАЛИЗ ГИПЕРССЫЛОК.

The paper is dedicated to cluster analysis of external web sites of large universities (web sites that refer to universities and web sites that are referred by universities). Web sites in Russia, the USA and the UK that have highest webometric ranking in their region were chosen as the subject of the study. The goal of the research is to identify a group of sites for each university that have the same kind of activity. The found clusters have been analyzed to determine the impact of group size and the number of groups on webometric ranking of university sites. To achieve the goal of the research, the authors developed a clustering algorithm based on the probabilistic method of reducing the dimension of multidimensional data (Locality-Sensitive Hashing, or LSH). An experiment that was conducted using the test data showed that the developed algorithm has good clustering quality and fast speed performance during massive dataset mining. The main results of the research are presented.

WEBOMETRICS; EXTERNAL WEB SITES OF UNIVERSITIES; CLUSTERING; LOCALITY-SENSITIVE HASHING; MIN HASHING; EXTERNAL WEB SITES CLUSTERING; HYPERLINKS ANALYSIS.

Для большинства крупных организаций немаловажное значение имеет их рейтинг, который рассчитывается в зависимости от

параметров, связанных с их родом деятельности. В частности, на общий рейтинг вуза большое влияние оказывает его вебметри-

* Данная статья является расширенной версией работы «Hierarchical Clustering of Large Text Datasets Using Locality-Sensitive Hashing» (авторы Vasilii Korelin, Ivan Blekanov), представленной на конференции International Workshop on Applications in Information Technology (IWAIT-2015) и опубликованной в материалах конференции: Puyshkin E.,

and Klyuev V. (Eds). Proceedings of the International Workshop on Applications in Information Technology (IWAIT-2015), The University of Aizu Press, 2015. Электронная версия доступна на сайте <<http://kspt.ftk.spbstu.ru/media/files/2015/iwait-2015/proceedings/iwait-2015-e-proceedings-release.pdf>>

ческий рейтинг. Известно, что одним из главных показателей, влияющих на вебометрический рейтинг любой организации, в т. ч. и университета, является количество внешних ресурсов [1], ссылающихся на сайт университета. Но кроме количества ссылающихся ресурсов также важно понять качество этих ресурсов, их природу, определить к какой области относится тот или иной внешний ресурс. Данные веб-ресурсы образуют группы сайтов с одинаковым родом деятельности. Таким образом, возникает задача выявления этих групп – кластеризации. Требуется определить степень влияния количества и размеров найденных групп на вебометрический рейтинг сайтов университетов. По найденным кластерам можно определить, с какими группами внешних ресурсов следует сайтам университетов выстраивать гиперссылочные взаимосвязи для повышения цитируемости.

Объектом данного исследования являются университетские сайты, имеющие чрезвычайно большие размеры (например, сайт СПбГУ содержит более 50 тыс. внутренних веб-страниц и около 5 млн гиперссылок) [2, 3]. Окружение таких сайтов может составлять десятки–сотни тысяч страниц. Стандартными методами кластеризации такого объема внешних веб-ресурсов не обойтись, т. к. при работе с большими коллекциями документов многие из данных методов показывают крайне неудовлетворительные результаты в плане производительности [4] (например, метод Single Linkage позволяет создавать кластеры произвольной формы, однако имеет высокую трудоемкость – $O(n^2)$, где n – число документов). Для того чтобы избавиться от «проклятия» размерности применяется вероятностный метод понижения размерности многомерных данных Locality-Sensitive Hashing – LSH, основная идея которого состоит в подборе хэш-функций для некоторых измерений для того, чтобы похожие объекты попадали в одну корзину [5].

Уменьшение размерности для анализа больших коллекций текстовых документов

Для преобразования текстовых документов в числовые множества в данной работе

использовался метод Shingling [6], который разбивает каждый документ на небольшие множества по k слов в каждом. Далее для сравнения документов применялась технология Min Hashing, позволяющая быстро сравнивать множества, содержащие большое число элементов.

Мера Жаккара. В работе для определения схожести двух текстов применяется мера Жаккара, определяющая схожесть множеств отношением числа элементов, входящих в пересечение двух множеств к числу элементов, входящих в объединение этих множеств [7]:

$$\text{Sim}(C_1, C_2) = \frac{|C_1 \cap C_2|}{|C_1 \cup C_2|}.$$

Так как каждый документ преобразуется во множества, состоящие из k слов, то набор документов можно представить в виде сильно разреженной булевой матрицы, где столбцы представляют собой документы, а строки – элементы универсального множества (например, множество элементов, где каждый элемент представляет собой множество k слов). Элемент такой матрицы равен единице, если документ (столбец) содержит данное множество k слов. В противном случае элемент матрицы равен нулю.

Minhashing. Для большой коллекции документов булева матрица будет сильно разрежена. Соответственно, матрица, описывающая такую коллекцию, займет много места в памяти. Кроме того, дальнейшая ее обработка также займет большое количество времени. Для того чтобы решить возникшие проблемы, данную матрицу преобразуем в матрицу, хранящую определенное количество хэш-функций и информацию о сходстве между похожими документами.

Minhash-функция $h(C)$ – это номер первой строки для столбца C в булевой матрице, где строки перемешаны случайным образом [8].

Как видим, число случайных перестановок задает число Minhash-функций. Например, можно использовать сто случайных перестановок для создания ста сигнатур для каждого столбца матрицы.

Сигнатуры могут быть записаны в другой матрице сигнатур (Signature matrix), чьи

колонки представляют собой документы, а строки – Minhash-значения.

Чем больше хэш-функций, тем выше вероятность того, что $\text{Sim}(C_1, C_2) = \text{Sim}(M[C_1], M[C_2])$, где C_i – столбец булевой матрицы, $M[C_i]$ – столбец матрицы сигнатур. Это важное свойство позволяет преобразовывать булевы матрицы с большим количеством строк в небольшие матрицы сигнатур, сохраняя сходство похожих множеств. Соответственно, похожесть двух столбцов определяется долей строк, в которых они равны.

Таким образом, каждый документ может быть представлен в виде вектора, число элементов которого равно количеству Minhash-функций.

Locality-Sensitive Hashing. Основная идея: сгенерировать из большого множества документов маленькие списки пар документов, чья схожесть должна быть посчитана. Для сравнения двух документов (столбцов) устанавливается порог t ($t < 1$). Пара документов считается похожей только в том случае, если доля одинаковых значений в матрице сигнатур больше t . Для матрицы сигнатур необходимо несколько раз вычислить хэш-значения и поместить документы с одинаковым значением в одну корзину (*bucket*). Документы, которые хоть раз попали в одну корзину, будут рассмотрены как кандидаты на сравнение [9].

Для множественного подсчета хэш-функций столбцов необходимо разбить матрицу сигнатур M на b частей по r строк в каждой. Для каждого b подсчитать хэш-значения столбцов и поместить столбцы с равным значением в одну корзину. Кандидатами на сравнение будут те столбцы, которые хоть раз попали в одну корзину. Для правильной работы алгоритма необходимо настроить r и b таким образом, чтобы похожие документы попадали в одну корзину, а непохожие – в разные.

Иерархическая кластеризация с использованием LSH

Данный алгоритм кластеризации использует хэш-таблицы, сформированные в результате LSH [10]. Данный алгоритм кластеризации с высокой долей вероятности создает такие же кластеры, как и в методе

Single linkage. Ниже представлено детальное описание алгоритма.

Предварительные условия:

$t < 1$ – порог, задающий минимальную схожесть документов;

$r = 1$ – начальное значение строк матрицы сигнатур в каждой группе;

r_{\min} – минимальное значение строк матрицы сигнатур в каждой группе;

Δ – коэффициент уменьшения параметра r .

Каждый документ представляет собой отдельный кластер.

Шаг 1. Для каждой группы b в каждом столбце вычислить хэш-функции и сохранить столбцы, у которых хэш-значения хотя бы раз попали в одну корзину, при условии, что в корзине должны находиться столбцы, принадлежащие разным кластерам. Если, например, какие-то два столбца принадлежат одному кластеру, то один из случайно выбранных столбцов удаляется из корзины.

Шаг 2. Для каждого из столбцов, входящих в одну корзину, отобрать пары кластеров, расстояние между которыми больше t .

Шаг 3. Пары кластеров, соответствующие парам, полученным на шаге 2, объединяются в один.

Шаг 4. Если $r \leq r_{\min}$, то алгоритм прекращает работу. Иначе, переход на шаг 5.

Шаг 5. $r = r - \Delta$. Переход на шаг 1.

Оценка качества кластеризации на основе LSH. Для оценки качества модифицированного метода агломеративной кластеризации с использованием LSH в данной работе использовалась тестовая коллекция текстовых документов Reuters-21578, содержащая 21 578 документов. На данной коллекции оценивались два метода иерархической кластеризации: стандартный алгоритм иерархической кластеризации Single-Link и модифицированный алгоритм Single-Link, основанный на применении LSH. В качестве основных метрик качества в работе использовались точность (R), полнота (P), аккуратность (Acc) и F -мера.

В табл. 1 приведены результаты оценки качества иерархической кластеризации для каждого из алгоритмов, а также время ра-

Таблица 1

Результаты оценки качества кластеризации на коллекции Reuters

Алгоритм	Количество документов	Acc, %	R, %	P, %	F-мера, %	Время работы
Single-Link	1000	75	79	60	68	4 с
	10 000	7	82	62	71	410 с
	20 000	—	—	—	—	> 1 ч
Single-Link + LSH	1000	72	81	66	73	3 с
	10 000	72	80	69	74	41 с
	20 000	78	83	64	72	90 с

боты каждого алгоритма при кластеризации 1000, 10 000 и 20 000 документов.

Как видно из таблицы, аккуратность и полнота метода Single-Link + LSH близки по значениям к аккуратности и полноте метода Single-Link. Точность разработанного метода иногда превосходит точность метода Single-Link. F-меры у обоих методов примерно одинаковые.

На рис. 1 приведена зависимость продолжительности работы алгоритмов кластеризации от числа входных документов.

Графики показывают, что алгоритм кластеризации с использованием LSH работает значительно быстрее, чем алгоритм без использования LSH. С увеличением числа

документов время работы алгоритма с использованием LSH линейно растёт.

Эксперимент. Исследование сайтов университетов методом кластеризации на основе LSH

Постановка эксперимента. В эксперименте требовалось для заданного списка сайтов университетов России, США и Великобритании, занимающих по своим регионам ведущие позиции в вебметрическом рейтинге [1], с помощью специализированного поискового робота [11] и базы данных Majestic [12] получить списки и содержимое всех внешних веб-страниц, которые их цитируют. А также с помощью

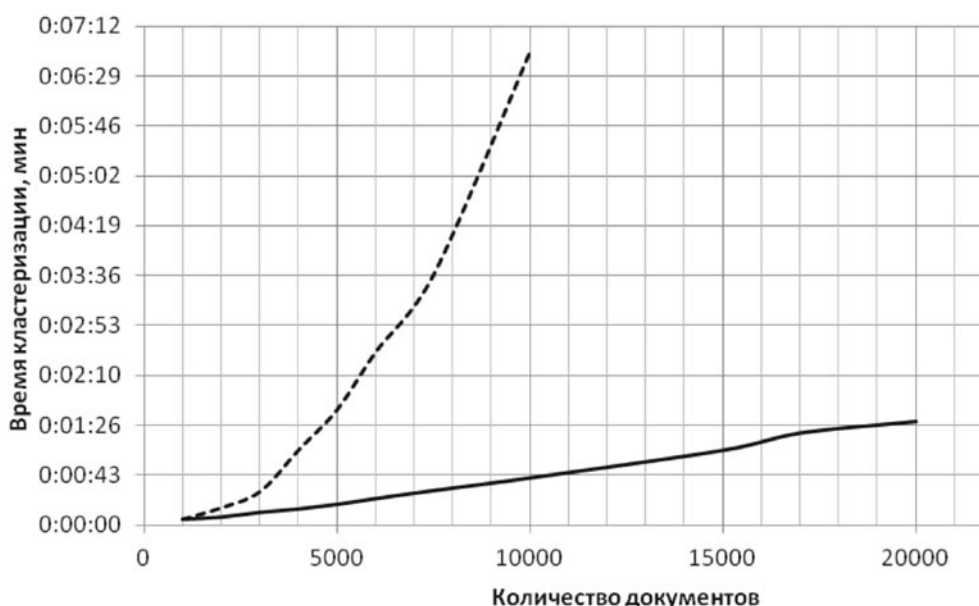


Рис. 1. Время работы алгоритмов кластеризации (---) Single-Link; (—) Single-Link + LSH

Таблица 2

Количество внешних ресурсов, окружающих университеты

URL-адрес вуза	Позиция в рейтинге Webometrics	Внешние ссылки на сайте вуза	Внешние домены на сайте вуза	Количество цитирующих сайт вуза веб-страниц	Количество цитирующих сайт вуза доменов
spbu.ru	539	3 280	469	1 599 059	16 028
msu.ru	129	817	262	7 039 127	39 416
nsu.ru	616	5 747	921	926 004	15 186
harvard.edu	1	918	163	75 994 723	319 445
stanford.edu	2	175	79	29 551 130	311 148
mit.edu	3	9	6	41 271 678	324 989
ox.ac.uk	16	8 075	1 631	8 920 524	117 959
cam.ac.uk	15	3 385	1 174	12 084 107	120 796
ucl.ac.uk	24	25 218	5 683	4 733 566	66 035

разработанного авторами метода агломеративной кластеризации на основе алгоритма LSH выявить целевые группы найденных внешних веб-ресурсов с одинаковым родом деятельности и установить степень их влияния на вебометрический рейтинг сайтов исследуемых вузов.

Для исследования были выбраны следующие сайты вузов:

Московский государственный университет имени М.В. Ломоносова (msu.ru), Россия;

Санкт-Петербургский государственный университет (spbu.ru), Россия;

Новосибирский государственный университет (nsu.ru), Россия;

Массачусетский технологический институт (mit.edu), США;

Гарвардский университет (harvard.edu), США;

Стэнфордский университет (stanford.edu), США;

Кембриджский университет (cam.ac.uk), Великобритания;

Оксфордский университет (ox.ac.uk), Великобритания;

Университетский колледж Лондона (ucl.ac.uk), Великобритания.

Результаты сбора. В итоге сбора данных

для анализа внешних веб-ресурсов сайтов исследуемых университетов были получены следующие результаты, представленные в табл. 2.

Таблица показывает, что университеты США, занимающие первые три позиции вебометрического рейтинга, имеют значительно больше внешних ресурсов, ссылающихся на них, чем ведущие университеты России и Великобритании. Ведущие университеты Великобритании также опережают российские по общему количеству ссылок и доменов, ссылающихся на них. Количество внешних ресурсов напрямую влияет на такой вебометрический индикатор, как Impact ресурса в Вебе.

На показатель видимости ресурса в Вебе, помимо числа внешних ссылающихся ресурсов, также влияет и их качество. Для кластеризации внешних веб-ресурсов были выбраны только англоязычные и русскоязычные страницы Веба. В табл. 3 приведено количество доменов для каждого университета, к которым применялся метод агломеративной кластеризации с использованием LSH.

Ниже приведены результаты кластеризации внешних ресурсов. Для определения их тематики извлекались наиболее часто

Таблица 3

Количество внешних ресурсов для кластеризации

Университет	Внешние домены на сайте вуза		Цитирующие сайт вуза домены	
	на русском	на английском	на русском	на английском
spbu.ru	342	81	7 373	3 929
msu.ru	177	62	3 343	1 008
nsu.ru	535	247	2 508	1 428
harvard.edu	4	151	1 545	16 632
stanford.edu	5	65	409	6 088
mit.edu	3	3	1 625	24 930
ox.ac.uk	29	1 399	126	3 187
cam.ac.uk	91	4 668	671	9 196
ucl.ac.uk	14	1 054	162	3 112

встречающиеся множества слов (shingles), которые присутствовали в кластере. В результате анализа были выделены основные часто встречающиеся группы: университеты, научные сообщества, поисковые системы, социальные сети (включая различные блоги, форумы и т. д.), медиасфера, новостные порталы и сайты других организаций.

Таким образом, определялась численность каждой группы для университета.

Результаты кластеризации внешних доменов на сайте вуза. Внешние домены – это домены, на которые ссылаются исследуемые сайты университетов. Ввиду ограниченного числа внешних ссылок на сайтах университетов США, рассматривались только

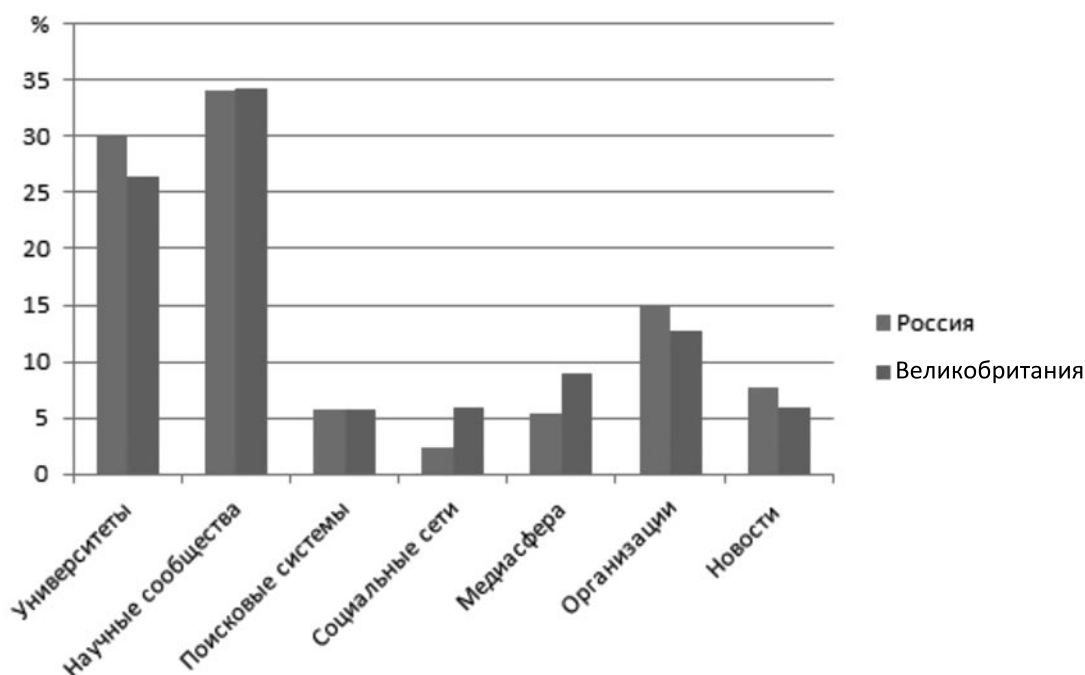


Рис. 2. Сравнительная гистограмма университетов России и Великобритании

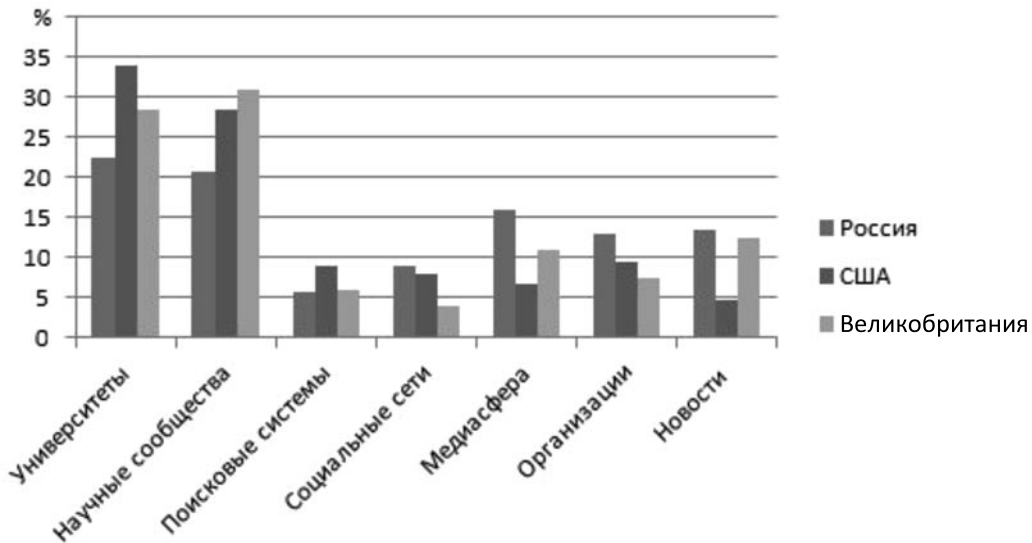


Рис. 3. Сравнительная гистограмма ресурсов, ссылающихся на университеты России, США и Великобритании

внешние ресурсы университетов России и Великобритании. Были получены кластеры для каждого из исследуемых университетов. Ниже представлена сравнительная гистограмма внешнего окружения университетов России и Великобритании (рис. 2).

Рисунок показывает, что внешние ресурсы, на которые ссылаются сайты университетов, практически совпадают. Университеты и научные сообщества составляют значительную долю внешних ресурсов, на которые ссылаются сайты университетов России и Великобритании.

Результаты кластеризации цитирующих сайт вуза доменов. Цитирующие веб-ресурсы – это сайты, сгруппированные по доменам, которые ссылаются на исследуемые сайты университетов. Были получены кластеры для каждого из исследуемых университетов. На рис. 3 приведена сравнительная гистограмма внешнего окружения университетов, представляющих Россию, США и Великобританию.

Анализ сайтов, ссылающихся на сайты вузов, показал, что университеты из одной страны имеют схожее внешнее окружение во многих областях. У всех университетов преобладают такие группы, как «Университеты» и «Научные сообщества». У университетов США и Великобритании эта доля выше, чем у российских. Соответственно,

в других областях доля российских университетов больше.

Можно сделать вывод о том, что чем выше доля университетов и научных сообществ среди внешних ресурсов университета, тем выше его вебметрический показатель Impact. Этот показатель зависит от количества и качества внешних ресурсов, ссылающихся на сайт. Внешние группы, содержащие университеты и научные сообщества, имеют значительно больший вес для вебметрического рейтинга, чем другие группы внешних ресурсов.

Разработан алгоритм иерархической кластеризации с использованием вероятностного метода понижения размерности многомерных данных Locality-Sensitive Hashing, оптимально подходящий для кластеризации больших коллекций документов (massive datasets). Данный алгоритм апробирован на тестовой коллекции текстовых документов Reuters. На тестовой коллекции алгоритм показал приемлемую точность (accuracy) и F -measure в сравнении с классическими методами иерархической кластеризации. Метод кластеризации с использованием LSH значительно превосходит по скорости работы классические методы иерархической кластеризации.

Разработанный алгоритм использовался для анализа веб-пространства нескольких университетов с их окружением.

Получены данные о внешнем окружении веб-пространства университетов Рос-

сии и Великобритании и о ресурсах, ссылающихся на сайты университетов России, США и Великобритании.

Работа выполнена при финансовой поддержке РФФИ, грант № 15-01-06105.

СПИСОК ЛИТЕРАТУРЫ

1. Ranking Web of Universities [электронный ресурс]/ URL: www.webometrics.info (дата обращения: 10.11.2015).

2. **Блеканов И.С., Москалец Р.Ю.** Теоретико-графовые характеристики в вебметрических исследованиях внутренней топологии крупных сегментов Веба // Процессы управления и устойчивость. 2015. Т. 2(18). № 1. С. 458–463.

3. **Blekanov I.S., Sergeev S.L., Maksimov A.I.** Analysis of the topology of large Web segments using Broder's bow-tie model // *Life Science Journal*. 2014. Vol. 11. Pp. 258–261.

4. **Ene A., Im S., Moseley B.** Fast clustering using MapReduce // In *KDD*. 2011. Pp. 681–689.

5. **Buhler J.** Efficient large-scale sequence comparison by locality-sensitive hashing // *Bioinformatics*. 2001. No. 17(5). Pp. 419–428.

6. **Broder A.Z.** Identifying and Filtering Near-Duplicate Documents // *Proc. of the 11th Annual Symp. on Combinatorial Pattern Matching*. 2000. Pp. 1–10.

7. **Jatsada Singthongchai, Suphakit Niwattanakul** A Method for Measuring Keywords Similarity by Applying Jaccard's, N-Gram and Vector Space

// *Lecture Notes on Information Theory*. Vol. 1. No. 4. Pp. 159–164.

8. **Chum O., Perdoch M., Matas J.** Geometric minhashing: Finding a (thick) needle in a haystack // *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2009. Pp. 17–24.

9. **Gionis A., Indyk P., Motwani R.** Similarity search in high dimensions via hashing // In *VLDB*. 1999. Pp. 518–529.

10. **Koga H., Ishibashi T., Watanabe T.** Fast agglomerative hierarchical clustering algorithm using Locality-Sensitive Hashing // *Knowledge and Information Systems*. 2007. Vol. 12. Iss. 1. Pp. 25–53.

11. **Блеканов И.С., Сергеев С.Л., Мартыненко И.А.** Построение тематико-ориентированных веб-краулеров с использованием обобщенного ядра // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. СПб.: Изд-во Политехн. ун-та, 2012. № 5(157). С. 9–15.

12. Marketing Search Engine and SEO Backlink Checker [электронный ресурс]/ URL: www.majestic.com (дата обращения: 02.04.2015).

REFERENCES

1. *Ranking Web of Universities*. Available: www.webometrics.info (Accessed: 10.11.2015).

2. **Blekanov I.S., Moskalets R.Yu.** Teoretiko-grafovyie kharakteristiki v vebometriceskikh issledovaniyakh vnutrenney topologii krupnykh segmentov Veba [Graph-theoretic characteristics of webometric studies internal topology of large segments of the Web]. *Protsessy upravleniya i ustoychivost* [Stability and Control Processes], 2015, Vol. 2(18), No. 1, Pp. 458–463. (rus)

3. **Blekanov I.S., Sergeev S.L., Maksimov A.I.** Analysis of the topology of large Web segments using Broder's bow-tie model. *Life Science Journal*, 2014, Vol. 11, Pp. 258–261.

4. **Ene A., Im S., Moseley B.** Fast clustering using MapReduce. In *KDD*, 2011, Pp. 681–689.

5. **Buhler J.** Efficient large-scale sequence comparison by locality-sensitive hashing. *Bioinformatics*, 2001, No. 17(5), Pp. 419–428.

6. **Broder A.Z.** Identifying and Filtering Near-Duplicate Documents. In *Proceedings of the 11th Annual Symposium on Combinatorial Pattern Matching*, 2000, Pp. 1–10.

7. **Jatsada Singthongchai, Suphakit Niwattanakul** A Method for Measuring Keywords Similarity by Applying Jaccard's, N-Gram and Vector Space. *Lecture Notes on Information Theory*, 2013, Vol. 1, No. 4, Pp. 159–164.

8. **Chum O., Perdoch M., Matas J.** Geometric minhashing: Finding a (thick) needle in a haystack. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, Pp. 17–24.

9. **Gionis A., Indyk P., Motwani R.** Similarity search in high dimensions via hashing. In *VLDB*, 1999, Pp. 518–529.

10. **Koga H., Ishibashi T., Watanabe T.** Fast agglomerative hierarchical clustering algorithm using Locality-Sensitive Hashing. *Knowledge and Information Systems*, 2007, Vol. 12, Issue 1, Pp. 25–53.

11. **Blekanov I.S., Sergeev S.L., Martynenko I.A.** Postroyeniye tematiko-oriyentirovannykh veb-kraulerov s ispolzovaniyem obobshchennogo yadra [The construction of the focused web-crawler using the universal kernel]. *Nauchno-tehnicheskie vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravlenie* [St. Petersburg State Polytechnical University



Journal. Computer Science. Telecommunications and Control System], St. Petersburg: SPbGPU Publ., 2012, No. 5(157), Pp. 9–15. (rus)

12. *Marketing Search Engine and SEO Backlink Checker*. Available: www.majestic.com (Accessed: 02.04.2015).

КОРЕЛИН Василий Николаевич – аспирант кафедры технологии программирования Санкт-Петербургского государственного университета.

199034, Россия, Санкт-Петербург, Университетская наб., д. 7-9.

E-mail: vn.korelin@gmail.com

KORELIN Vasilii N. *St. Petersburg State University.*

199034, Universitetskaya Emb. 7-9, St. Petersburg, Russia.

E-mail: vn.korelin@gmail.com

БЛЕКАНОВ Иван Станиславович – доцент кафедры технологии программирования Санкт-Петербургского государственного университета, кандидат технических наук.

199034, Россия, Санкт-Петербург, Университетская наб., д. 7-9.

E-mail: i.blekanov@gmail.com

BLEKANOV Ivan S. *St. Petersburg State University.*

199034, Universitetskaya Emb. 7-9, St. Petersburg, Russia.

E-mail: i.blekanov@gmail.com

СЕРГЕЕВ Сергей Львович – заведующий кафедрой технологии программирования Санкт-Петербургского государственного университета, кандидат физико-математических наук, доцент.

199034, Россия, Санкт-Петербург, Университетская наб., д. 7-9.

E-mail: slsergeev@yandex.ru

SERGEEV Sergei L. *St. Petersburg State University.*

199034, Universitetskaya Emb. 7-9, St. Petersburg, Russia.

E-mail: slsergeev@yandex.ru

DOI: 10.5862/JCSTCS.229.9

UDC 004.627

E.M. Khassina, A.A. Lomov

AUDIO FILES COMPRESSION WITH THE STLS-ESM METHOD*

E.M. Хасина, А.А. Ломов

СЖАТИЕ АУДИОФАЙЛОВ МЕТОДОМ STLS-ESM

In the paper an audio compression algorithm based on modeling an audio signal by a partial solution of a certain difference equation in the time domain is investigated. The signal is modeled as a sum of exponentially damped sinusoids. Such an approach is thought to be efficient in modeling the transient segments that are present in speech audio signals and audio signals generated by conventional musical instruments. In order to approximate an audio signal frame with a solution of a difference equation, a variational (STLS) problem is solved using the inverse iteration algorithm with an updating inverse matrix. The α -version of the audio codec based on the STLS-ESM scheme was created and tested in comparison with LAME MP3 codec.

AUDIO SIGNALS MODELING; AUDIO CODEC; EXPONENTIAL SINUSOIDAL MODEL; PARAMETRIC IDENTIFICATION; DIFFERENCE EQUATIONS; VARIATIONAL IDENTIFICATION METHOD; STRUCTURED TOTAL LEAST SQUARES.

Рассмотрен алгоритм сжатия аудиосигнала путем его аппроксимации во временной области частным решением некоторого разностного уравнения. Аудиосигнал представлен в виде суммы экспоненциально затухающих/растущих синусоид. Такой подход эффективен для моделирования переходных процессов в речевых сигналах и сигналах традиционных музыкальных инструментов. Для построения аппроксимации решена вариационная задача идентификации с использованием итераций с обновляемой обратной матрицей. Представлена α -версия аудиокодека и приведены результаты тестирования в сравнении с аудиокодеком LAME MP3.

МОДЕЛИРОВАНИЕ АУДИОСИГНАЛОВ; АУДИОКОДЕК; EXPONENTIAL SINUSOIDAL MODEL; ПАРАМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ; РАЗНОСТНЫЕ УРАВНЕНИЯ; МЕТОД ВАРИАЦИОННОЙ ИДЕНТИФИКАЦИИ.

1. INTRODUCTION

Lossless audio codecs, e. g. FLAC, Monkey's Audio APE, TTA can achieve compression ratios of 2-3 times.

Lossy audio codecs such as MPEG-1 codecs (MP1, MP2 and MP3) and OGG

Vorbis reduce the sound quality of music files, without, however, degrading it radically. Hence, for the human ear, the distortions made by such codecs are not actually noticeable. A music file compressed to a fifth of its original size still has the sound quality of radio broadcasting.

In their work, lossy codecs tend to decompose an audio signal into harmonics. However, the method often does not correspond to the physical nature of sounds produced by conventional musical instruments, for which the presence of a considerable quantity of transients (high-amplitude, short-duration sound segments followed by an exponential decay) is common. If a piece of a signal contains transients it cannot be considered as a quasi-stationary episode. This is why the MP3 audio codec, for instance, has to use the

*This article is an extended post-conference revision of the paper «Audio Files Compression with the Variational Method of Identification of Modeling Difference Equations» by Eugenia M. Khassina and Andrei A. Lomov published in Pyshkin, E., and Klyuev, V. (Eds). Proceedings of the International Workshop on Applications in Information Technology (IWAIT-2015), The University of Aizu Press, 2015. Available at <<http://kspt.ftk.spbstu.ru/media/files/2015/iwait-2015/proceedings/iwait-2015-e-proceedings-release.pdf>>

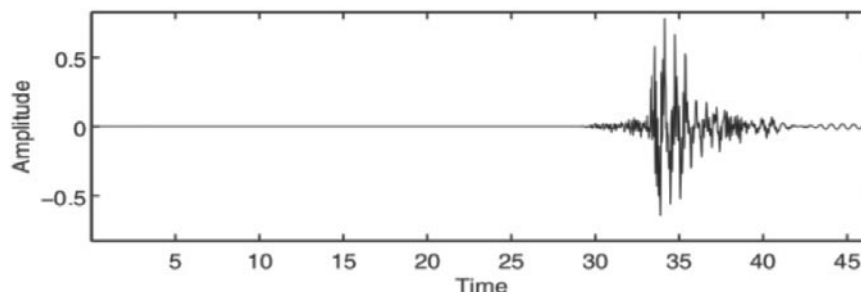


Fig. 1. Transient [1]

Modified Discrete Cosine Transform (MDCT) with windows of varied length while processing a signal. A long window (1024-subband MDCT) is used for quasi-stationary episodes of a signal while a short window (128-subband MDCT) is used for episodes with transient attacks. Providing a better time resolution, the short-block mode of the encoding scheme also helps to avoid what is commonly referred to as a pre-echo artifact: frequently annoying audible quantization noise that occurs before a transient [1].

In Fig. 1 a transient can be seen. Even though it is more natural to decompose such signals into a sum of exponentially damped sinusoids, rather than to represent them as a Fourier series, such an approach requires a considerable amount of CPU resources. This is the reason why the creation and the usage of codecs based on the principle have only recently become justified.

In formula (1) below, audio signal $s(n)$ is represented as a superposition of slowly time-varying exponentially weighted sinusoids and quasi-stationary noise $\eta(n)$. The signal model is called the Exponential Sinusoidal Model (ESM). The frequencies ω_i , phases ϕ_i , amplitudes a_i and damping parameters γ_i can be obtained without switching to the frequency domain.

$$s(n) \approx \sum_{i=1}^K a_i(n) e^{-\gamma_i(n)n} \sin(\omega_i(n)n + \phi_i(n)) + \eta(n). \quad (1)$$

In Ref. [2] a Total Least Squares (TLS) problem of order $2K$ is solved to estimate ω_i and γ_i , where K is the number of available exponentially damped sinusoids predefined by a user [3, 4]. On the basis of the TLS-ESM scheme an experimental audio codec was created and tested [2].

TLS searches for the parameters of the damped sinusoids maximizing the signal-to-noise ratio (SNR), that is, the algorithm tries to make the original and the modeled signals as close to each other as possible in the time domain on each set of $2K$ samples. TLS considers the sets independent even if they contain common samples, and if the sets are dependent in reality. The problem is an essential disadvantage of TLS.

It also should be noted that the SNR quality measure is a formal criterion. It effectively allows to achieve a high degree of time resemblance between the original signal and the modeled one on separate sets of samples while ignoring the perceptual quality of the modeled audio that is much more important for an audio codec.

Experiments showed [2] that TLS spends all the modeling sinusoids for low frequencies with high amplitudes. An explanation for this fact is that adding a high-amplitude sinusoid to the modeled signal reduces the difference between the original and the modeled signals more than adding a high-frequency low-amplitude sinusoid does. As a result, there are few available sinusoids left for modeling the high frequencies and this is quite noticeable for the human ear.

Thus, the initial fullband TLS algorithm had to be improved with additional signal processing in the frequency domain. The modified version of the algorithm is called Subband TLS-ESM [2]. The original signal is decomposed into 32 frequency subbands with a fully decimated uniform QMF analysis filter bank (32 channels) used in the MPEG-1 Layer I codec. For each of the subband signals an independent TLS problem is solved with the same number of sinusoids employed for the modeling.

The goal of the work is to research an audio compression algorithm which decomposes a signal into exponentially damped sinusoids in the time domain without switching to the frequency domain. We use an approximation of an audio signal on the whole audio frame using the variational identification method [4–6], close to the Structured Total Least Squares (STLS) [7] and the Global Total Least Squares (GTLS) [8] methods.

In [9] a vocoder based on the ESM-STLS scheme was proposed. The testing of the vocoder was performed in comparison to Code-excited linear prediction (CELP), a standard speech coding algorithm. The results of the testing showed that, providing a similar compression ratio, the new vocoder has a substantially higher signal-to-noise ratio (SNR). However, the speech spectrum is rather simple, which makes speech signals easily compressible. Our experiments showed that Newton’s iterative algorithms, to which STLS1 and STLS2 used in [9] belong, have bad convergence when solving the STLS problem for music audio files with wider spectra.

2. THEORETICAL ASPECTS

Coding an audio frame. Our algorithm divides the whole signal of an audio file into frames of N samples each, processing the frames one by one. In section 3 we will consider how N value and other parameters are chosen. Conventionally, N equals 100. A frame of samples is denoted by $s[k], k = \overline{1, N}$.

We will treat the vector $s \doteq (s[1]; \dots ; s[N])$ as a perturbed observation of a solution process $z \doteq (z[1]; \dots ; z[N])$ of a certain homogeneous linear difference equation with real coefficients. We will solve the inverse problem of identifying the unknown coefficients of the equation. Let us take as an example a difference equation of order $p = 3$:

$$z[k + 3] + \alpha_2 z[k + 2] + \alpha_1 z[k + 1] + \alpha_0 z[k] = 0, \quad k = \overline{1, N - 3}. \quad (2)$$

Let us denote the characteristic roots of the system (2) by ξ_i . The characteristic polynomial of the system is real, hence all the roots that are complex should occur in complex-conjugate pairs. For the present example, let us suppose

that the single real root of the characteristic polynomial is ξ_2 . Then we have the general solution of (2) as follows:

$$z[k] = C_0 \xi_1^k + C_1 \overline{\xi_1}^k + C_2 \xi_2^k. \quad (3)$$

We are interested only in real solutions of the difference equation, as audio samples of observation s are real. Therefore, we can switch to the real form of the general solution (3), assuming C_0 and C_1 to be complex conjugates, and convert it to a sum of exponentials and exponentially damped sinusoids:

$$\begin{aligned} z[k] &= C_1 \xi_1^k + \overline{C_1} \overline{\xi_1}^k + C_2 \xi_2^k = \\ &= C_1 e^{(\gamma_1 + i\omega_1)k} + \overline{C_1} e^{(\gamma_1 - i\omega_1)k} + C_2 e^{\gamma_2 k} = \\ &= A_1 \rho_1^k \cos(k\omega_1) + A_2 \rho_1^k \sin(k\omega_1) + \\ &\quad + A_3 \rho_2^k, \quad \forall i \ A_i \in \mathbb{R}. \end{aligned} \quad (4)$$

We introduce the following notation for the vector of the coefficients of the difference equation:

$$\gamma \doteq (\alpha_0 \quad \alpha_1 \quad \alpha_2 \quad 1)^T.$$

Let us define the objective function for the identification of vector γ :

$$\begin{aligned} J(\gamma) &= \|s - z(\gamma)\|^2, \\ z(\gamma) &\doteq \arg \min_{z: (2)} \|s - z\|^2. \end{aligned} \quad (5)$$

This variational problem was first formulated and solved by A.O. Egorshin [4, 5]. For its numerical solution we apply the iterative algorithm with an updating inverse matrix proposed by A.O. Egorshin and, independently, by M.R. Osborne [10]. We use the least-squares estimate γ_{LS} [11, 12] as the initial γ for the iterative algorithm w. Now we define special matrices in order to derive the formula for γ_{LS} and to describe the iterative algorithm solving problem (5).

First, let us transform the difference equation (2) to a matrix form:

$$\underbrace{\begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & 1 & & 0 \\ 0 & \alpha_0 & \alpha_1 & \alpha_2 & 1 & \\ \vdots & & \ddots & \ddots & \ddots & \ddots \\ 0 & & & \alpha_0 & \alpha_1 & \alpha_2 & 1 \end{pmatrix}}_G \underbrace{\begin{pmatrix} z[1] \\ z[2] \\ \vdots \\ z[N] \end{pmatrix}}_z = 0. \quad (6)$$

We search for vector γ by adjusting the residual $e_\gamma \doteq G_\gamma s$ to approach zero. Now the objective function for the least-squares problem is as follows:

$$J_\gamma = \arg \min_\gamma e_\gamma^T e_\gamma, \quad e_\gamma = G_\gamma s. \quad (7)$$

Then we use the identity $G_\gamma s \equiv V(s)\gamma$, where V is a Hankel matrix:

$$\underbrace{\begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & 1 & & 0 \\ 0 & \alpha_0 & \alpha_1 & \alpha_2 & 1 & \\ \vdots & & \ddots & \ddots & \ddots & \ddots \\ 0 & & & \alpha_0 & \alpha_1 & \alpha_2 & 1 \end{pmatrix}}_{G_\gamma} \underbrace{\begin{pmatrix} s[1] \\ s[2] \\ \vdots \\ s[N] \end{pmatrix}}_s \equiv \underbrace{\begin{pmatrix} s[1] & s[2] & s[3] & s[4] \\ s[2] & s[3] & s[4] & s[5] \\ \vdots & \vdots & \vdots & \vdots \\ s[N-3] & s[N-2] & s[N-1] & s[N] \end{pmatrix}}_{V_1} \underbrace{\begin{pmatrix} s[4] \\ s[5] \\ \vdots \\ s[N] \end{pmatrix}}_{V_2} \underbrace{\begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ 1 \end{pmatrix}}_\gamma. \quad (8)$$

We denote the vector of the identified coefficients by θ and rewrite the expression for the residual e_γ :

$$\theta = (\alpha_0 \quad \alpha_1 \quad \alpha_2)^\top, \quad \gamma = \begin{pmatrix} \theta \\ 1 \end{pmatrix}, \quad (9)$$

$$e_\gamma = V(s) \cdot \gamma = V_1(s) \cdot \theta + V_2(s).$$

Solving the linear least-squares problem we get the values of the coefficients θ :

$$\theta_{LS} = -(V_1^\top V_1)^{-1} V_1^\top V_2. \quad (10)$$

Now, the iterations with an updating inverse matrix which solve the variational identification problem (5) are:

1. The initial value: $\gamma = \gamma(0) = \gamma_{LS}$.
2. For $k \geq 0$

$$\begin{cases} \tau = (V(s)^\top (G_{\gamma(k)} G_{\gamma(k)}^\top)^{-1} V(s))^{-1} \cdot \gamma(k), \\ \gamma(k+1) = \frac{1}{(0\dots 01)\tau} \tau. \end{cases} \quad (11)$$

The last equation means the division of the whole auxiliary vector θ by its last element in order to make the last element of vector $\gamma(k+1)$ equal to unity. The main difference of the Egorshin–Osborne iterations from the computational TLS algorithm consists in the presence of the inverse matrix $(G_{\gamma(k)} G_{\gamma(k)}^\top)^{-1}$ which is updated at each iteration.

Using the calculated estimate for γ , we find the modeling process $z(\gamma)$ nearest to the observation s as the linear projection [4, 5]:

$$z(\gamma) = \underbrace{(I - G_\gamma^\top (G_\gamma G_\gamma^\top)^{-1} G_\gamma)}_{I - \Pi} \cdot s, \quad (12)$$

where I is an identity matrix.

Matrix $I - \Pi$ is a projection matrix to the subspace of solutions of the equation $G_\gamma z = 0$.

The obtained coefficient vector γ and the corresponding process $z(\gamma)$ that fit the observation s are used further, as we will show below.

Note also that in this section the order p of the difference equation is considered known. A way of choosing p and the problems encountered when using the iterations (11) at the implementation stage will be conveyed in the next section.

Decoding an audio frame. From (4) we can see that, in order to restore process z , for each complex-conjugate pair of roots of the characteristic polynomial of the difference equation we need to know the real argument and the real modulus of the polar form of that root of the pair that lies above the real axis and we also need to know the set of real coefficients A_i . Thus, we should keep $2p$ float numbers to restore one audio frame. Besides, we should also keep one byte (or two/three for the last frame of an audio file) of service information for each audio frame such as the order of a difference equation and the frame size type. We will not describe in detail the structure of the frame service information byte in this paper.

The only thing left for us to understand is how to get coefficients $A_i, i = 1, p$. Let us transform the expression (4) to a matrix form:

$$\begin{pmatrix} z[1] \\ \vdots \\ z[N] \end{pmatrix} = \begin{pmatrix} \rho_1 \cos(\omega_1) & \rho_1 \sin(\omega_1) & \rho_2 \\ \rho_1^2 \cos(2\omega_1) & \rho_1^2 \sin(2\omega_1) & \rho_2^2 \\ \vdots & \vdots & \vdots \\ \rho_1^N \cos(N\omega_1) & \rho_1^N \sin(N\omega_1) & \rho_2^N \end{pmatrix} \times \quad (13)$$

$$\times \underbrace{\begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}}_d = Hd. \quad (13)$$

Using the iterations (11) we have found the model G and the process z , corresponding to the original observation s , such that $Gz = 0$. Knowing G (the coefficients of the difference equation), we can find the characteristic roots of the equation and, thus, the matrix H . Note that the next expression is true:

$$Gz = GHd = 0, d \neq 0 \Rightarrow G \perp H.$$

The needed vector d containing coefficients A_i can be found with the least squares method:

$$d = (\bar{H}^T \bar{H})^{-1} \bar{H}^T \bar{z}, \quad (14)$$

where \bar{H} is a submatrix composed of $\geq p$ rows of matrix H and \bar{z} is a subvector composed of the corresponding elements of vector z . Also the following equations take place:

$$\begin{aligned} d &= (H^T H)^{-1} H^T z = \\ &= (H^T H)^{-1} H^T H (H^T H)^{-1} H^T s = \\ &= (H^T H)^{-1} H^T s. \end{aligned}$$

3. CODEC IMPLEMENTATION

We have implemented an audio codec (consisting of the coder and decoder modules), based on the theory described above, in the

Scilab cross-platform environment (<http://www.scilab.org>) supplying a high-level programming language resembling MATLAB, the language interpreter and an ample set of mathematical functions. The codec implementation and testing were performed in the Debian GNU/Linux operating system. As input the codec accepts a mono WAV audio file with a sample rate of 44.1 KHz and a bit depth of 16 bits. As output the codec produces a compressed file, whose special structure we had to define with a new extension `.cod`. In Fig. 2 you can see the general scheme of the codec operation.

An original audio signal to be compressed is divided into frames of N samples with a shift of $(N - M)$ samples. That is to say, each pair of consequent frames overlap by M samples to be glued smoothly after their decompression. After the decoding procedure two neighboring frames are summed on the gluing area preliminarily weighted. The weighting coefficients vary from 0 to 1. The weighting functions we use are $\sin^2(ck)$ and $1 - \sin^2(ck)$ where the time index k runs from 0 to $M - 1$ and $c(M - 1) = \pi/2$. In Fig. 3 you can see the gluing scheme.

The values of N and M can be predefined by a user. It was discovered that for frames longer than 150 samples the coding procedure often fails because the iterations (11) do not converge, and if we take M value < 5 , an audible noise appears on joints of neighboring frames. We chose the average values: $N = 100$ and $M = 10$.

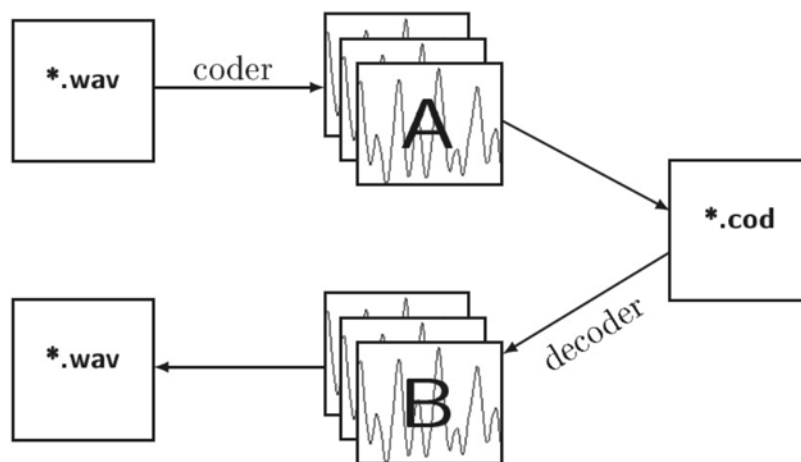


Fig. 2. General scheme of the codec operation:
A. Original frames. B. Decompressed frames

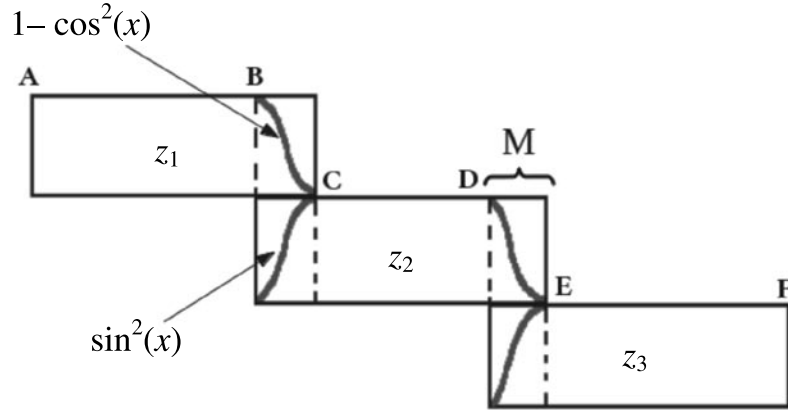


Fig. 3. Gluing of frames after decompression

As an example, let us have an initial audio signal s consisting of 280 samples. We will model it with three frames that consist of $N = 100$ samples each and overlap by $M = 10$ samples.

Let us denote the first frame we are to approximate by $s_1 = (s[1]; \dots; s[100])$. We find $2p$ float numbers needed to restore s_1 . We denote the restored version of the frame by z_1 and we will call its endings A and C as it is shown in Fig. 3.

Now we take the following frame $s_2 = (s[91]; \dots; s[100]; \dots; s[190])$ and model it. The result of the frame modeling is z_2 whose endings are B and E.

The two modeled frames overlap in the area [B, C]. We multiply the overlapping part ($z_1[91]; \dots; z_1[100]$) of z_1 by a decreasing square sinusoid

$$\left(\sin^2\left(\frac{\pi}{2}\right); \sin^2\left(\frac{\pi \cdot M - 2}{2 \cdot M - 1}\right); \dots; \right. \\ \left. \sin^2\left(\frac{\pi \cdot 1}{2 \cdot M - 1}\right); \sin^2(0) \right)$$

component-wise. The dimensions of the multiplied vectors are both equal to M . At the same time we multiply the overlapping part ($z_2[1]; \dots; z_2[10]$) of z_2 by the vector of an increasing square sinusoid

$$\left(\sin^2(0); \sin^2\left(\frac{\pi \cdot 1}{2 \cdot M - 1}\right); \dots; \right. \\ \left. \sin^2\left(\frac{\pi \cdot M - 2}{2 \cdot M - 1}\right); \sin^2\left(\frac{\pi}{2}\right) \right).$$

The dimensions of the vectors multiplied are equal again.

The result of gluing z_1 and z_2 is a consequence of the samples of [A, B] part of z_1 , the samples of a component-wise sum of the overlapping parts [B, C] of z_1 and z_2 and the samples of [C, E] part of z_2 . The resulting vector [A, E] consists of 190 samples.

Now we model the last frame s_3 . The result of the modeling is frame z_3 with the endings D and F. We glue the resulting vector [A, E] with z_3 in the same way. Now the overlapping area is [D, E].

The described approach was implemented in our codec and all the tests showed that the gluing is performed smoothly. However, we should note that the overlaps decrease the resulting compression ratio because the quantity of samples that we can code using n frames declines from $n \cdot N$ to $n \cdot (N - M) + M$ when the frames overlap.

When coding a frame we increment a model order p in a cycle from $p_{\min} = 2$ to $p_{\max} = 13$. For each p the identification of coefficients of the equation (2) is performed and the model process (12) is found. After this, the relative modeling error is counted as follows:

$$\frac{\|z - s\|}{\|s\|} \leq 5\%, \quad (15)$$

where the value 5 % is the relative error threshold. If the relative error counted does not exceed the threshold then we break the p cycle and write the found $2p$ float numbers and a service information byte to the output

Testing results

Files type	Files number	Our codec		LAME MP3 128 kbps		LAME MP3 256 kbps	
		compression ratio	relative error	compression ratio	relative error	compression ratio	relative error
Piano	20	3.334	0.028	5.15	0.052	2.575	0.001
Electric guitar	20	1.451	0.028	5.15	0.056	2.575	0.003

.cod file. Obviously, the least suitable model order is preferable to make the compressed file as small as possible. If the coder fails for any order p with the chosen relative precision 5 % then it divides the frame in half and tries to model each of the two smaller frames again. If the modeling process for a smaller frame is not successful anyway, then the frame is written to .cod file directly without compression. The relative error threshold is predefined by a user and, in general, can be set to an arbitrarily small number but it would lead to a low compression ratio.

4. TESTING

The testing was performed over 20 piano audio files and 20 electric guitar audio files of 44 100 samples each (one second duration) for our audio codec and also for LAME MP3 [version 3.99.5] [x86] codec with constant bit rates (CBR) 128 Kbps and 256 Kbps in order for us to be able to assess the effectiveness of our codec comparing to it. You can see the results of the testing in the Table.

We compressed each original WAV audio file with a coder to a .cod or MP3 file and then decoded the compressed file to a WAV file again. After that, we counted the relative error between the audio signal of the original WAV file and the signal of the decompressed WAV file as shown in (15). The relative error values presented in the Table are average for both sets of 20 audio files. The relative error threshold in our codec is 5 %, thus, the average relative error values relating to our codec are less than 0.05.

Using a bit rate of 128 Kbps typically results in a sound quality equivalent to what we would hear on the radio. As you can see the

relative error for our codec is less than the one for LAME MP3 codec with CBR 128 Kbps. However, the relative error is only an objective sound quality measurement. When we were assessing the subjective perceptual quality of the sound produced by our codec by listening to it in headphones, the sound happened to be distinctly worse than the sound of the audio files produced by LAME MP3 codec with CBR 128 Kbps.

The compression ratio values related to our codec are average for both sets of 20 audio files in the Table. The compression ratio reached by LAME MP3 was identical for all the audio files (5.15 times for 128 Kbps and 2.575 times for 256 Kbps) as we used it in the constant bit rate mode.

Considering the work of our codec, one can also notice that the average compression ratio reached by the codec for “simple” piano files is two times bigger than the ratio for “complicated” electric guitar files. The reason for this is that the iterations (11) converge worse for the latter. Therefore, more frames of an electric guitar file are written fully to an output compressed file .cod increasing its size.

We consider the codec as an interesting application of parametric identification methods in the time domain. The key point of its work is the variational (STLS) objective function (5) that is minimized in our modeling algorithm. The iterations (11) minimize the function over difference equation coefficients effectively for simple piano music files and the algorithms STLS1 and STLS2 used in [9] also solve the STLS problem well for speech signals. However, the STLS approach does not work properly for more complicated music files. We are going to



handle the problem by dividing an audio signal into frequency subbands and coding each of them independently. Additionally, we are planning to search for more efficient ways of minimizing the variational objective function; one option

is to do it over the roots of the characteristic polynomial of the difference equation.

The work has been supported by the Russian Foundation for Basic Research (project No. 13-01-00329).

REFERENCES

1. **You Y.** *Audio Coding Theory and Applications*, New York: Springer, 2010.
2. **Hermus K., Verhelst W., Lemmerling P., Wambacq P., van Huffel S.** Perceptual Audio Modeling with Exponentially Damped Sinusoids. *Signal Processing*, 2005, Vol. 85, No. 1, Pp. 163–176.
3. **de Groen P.P.N.** An Introduction to Total Least Squares. *Nieuw Archief voor Wiskunde*, 1996, Vol. 14, No. 4, Pp. 237–253.
4. **Lomov A.A.** Variational identification methods for linear dynamic systems and the local extrema problem. *Upravlenie Bol'shimi Sistemami*, 2012, Vol. 39, Pp. 53–94. Available: <http://ubs.mtas.ru/upload/library/UBS3903.pdf> (Accessed: 01.11.2015). (rus)
5. **Egorshin A.O.** *Computational closed algorithms of identification of linear objects. Optimal and Self-Adjusting Systems*, Novosibirsk, 1971, Pp. 40–53.
6. **Egorshin A.O.** Least square method and fast algorithms in variational problems of identification and filtration (VI method). *Avtometriya*, 1988, Vol. 1, Pp. 30–42. (rus)
7. **Moor B.D.** Structured total least squares and L2 approximation problems. *Linear Algebra and its Applications*, 1993, Vol. 188–189, Pp. 163–207.
8. **Roorda B., Heij C.** Global total least squares modelling of multivariable time series. *The IEEE Transactions on Automatic Control*, 1995, Vol. AC-40, Pp. 50–63.
9. **Lemmerling P., Mastronardi N., van Huffel S.** Efficient implementation of a structured total least squares based speech compression method. *Linear Algebra and its Applications*, 2003, Vol. 366, Pp. 295–315.
10. **Osborne M.R., Anderssen R.S.** *A class of nonlinear regression problems. Data Representation*, Saint Lucia: University of Queensland Press, 1970, Pp. 94–101.
11. **Khassina E.M.** Audio Files Compression with the Variational Method of Identification of Modeling Difference Equations. *Proceedings of the International Workshop on Applications in Information Technology*, The University of Aizu in cooperation with St. Petersburg State University and Peter the Great St. Petersburg Polytechnic University, Aizu-Wakamatsu, Japan, Oct. 2015, Pp. 1–4. Available: <http://kspt.ftk.spbstu.ru/media/files/2015/iwait-2015/proceedings/iwait-2015-e-proceedings-release.pdf> (Accessed: 01.11.2015).
12. **Khassina E.M.** File compression by use of the method of variation identification of modeling difference equations. *Proceedings of the X International Conference System Identification and Control Problems SICPRO'15*, Moscow, 2015, Pp. 648–658. Available: <http://www.sicpro.org/sicpro15/proc/procdngs/648.pdf> (Accessed: 01.11.2015).

СПИСОК ЛИТЕРАТУРЫ

1. **Ю. Ю.** *Audio Coding Theory and Applications*. New York: Springer, 2010.
2. **Гермус К., Верхелст В., Леммерлинг П., Вамбацк П., ван Хуффел С.** Perceptual Audio Modeling with Exponentially Damped Sinusoids // *Signal Processing*. 2005. Vol. 85. No. 1. Pp. 163–176.
3. **де Гроен П.П.Н.** An Introduction to Total Least Squares // *Nieuw Archief voor Wiskunde*. 1996. Vol. 14. No. 4. Pp. 237–253.
4. **Ломов А.А.** Вариационные методы идентификации линейных динамических систем и проблема локальных экстремумов // *Управление большими системами*. 2012. № 39. С. 53–94 [электронный ресурс] / URL: <http://ubs.mtas.ru/upload/library/UBS3903.pdf> (дата обращения: 01.11.2015).
5. **Егоршин А.О.** Computational closed algorithms of identification of linear objects. *Optimal and Self-Adjusting Systems*. Новосибирск, 1971. С. 40–53.
6. **Егоршин А.О.** Метод наименьших квадратов и быстрые алгоритмы в вариационных задачах идентификации и фильтрации (метод ВИ) // *Автометрия*. 1988. № 1. С. 30–42.
7. **Moor B.D.** Structured total least squares and L2 approximation problems // *Linear Algebra and its Applications*. 1993. Vol. 188–189. Pp. 163–207.
8. **Roorda B., Heij C.** Global total least squares

modelling of multivariable time series // The IEEE Transactions on Automatic Control. 1995. Vol. AC-40. Pp. 50–63.

9. **Lemmaerling P., Mastronardi N., van Huffel S.** Efficient implementation of a structured total least squares based speech compression method // Linear Algebra and its Applications. 2003. Vol. 366. Pp. 295–315.

10. **Osborne M.R., Anderssen R.S.** A class of nonlinear regression problems. Data Representation. Saint Lucia: University of Queensland Press, 1970. Pp. 94–101.

11. **Khassina E.M.** Audio Files Compression with the Variational Method of Identification of Modeling Difference Equations // Proceedings of the International Workshop on Applications in In-

formation Technology (IWAIT-2015). The University of Aizu in cooperation with St. Petersburg State University and Peter the Great St. Petersburg Polytechnic University. Aizu-Wakamatsu, Japan, 2015. Pp. 1–4 [электронный ресурс] / URL: <http://kspt.ftk.spbstu.ru/media/files/2015/iwait-2015/proceedings/iwait-2015-e-proceedings-release.pdf> (дата обращения: 01.11.2015).

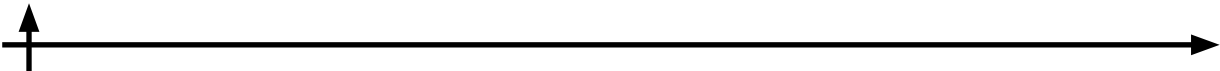
12. **Khassina E.M.** Сжатие аудиофайлов методом вариационной идентификации моделирующих разностных уравнений // Матер. X международн. конф. Идентификация систем и задачи управления. SICPRO-15. М., 2015. С. 648–658 [электронный ресурс] / URL: <http://www.sicpro.org/sicpro15/proc/procdngs/648.pdf> (дата обращения: 01.11.2015).

KHASSINA Eugenia M. *Novosibirsk State University.*
630090, Pirogov Str. 2a, Novosibirsk, Russia.
E-mail: jenua-100@yandex.ru

ХАССИНА Евгения Михайловна — магистрант кафедры систем информатики Новосибирского государственного университета.
630090, Россия, г. Новосибирск, ул. Пирогова, д. 2а.
E-mail: jenua-100@yandex.ru

LOMOV Andrei A. *Sobolev Institute of Mathematics.*
630090, Acad. Koptuyug Str. 4, Novosibirsk, Russia.
E-mail: lomov@math.nsc.ru

ЛОМОВ Андрей Александрович — старший научный сотрудник Института математики имени С.Л. Соболева СО РАН, доктор физико-математических наук.
630090, Россия, г. Новосибирск, ул. Коптюга, д. 4.
E-mail: lomov@math.nsc.ru



Инфокоммуникационные технологии в образовании

DOI: 10.5862/JCSTCS.229.10

УДК 004.7

М.Ю. Маслов, С.М. Носницын, А.В. Самочадин, К.Е. Логинов

АРХИТЕКТУРА СРЕДСТВ УПРАВЛЕНИЯ МОБИЛЬНЫМИ УСТРОЙСТВАМИ ДЛЯ ОБРАЗОВАТЕЛЬНЫХ УЧРЕЖДЕНИЙ

M.Yu. Maslov, S.M. Nosnitsyn, A.V. Samochadin, K.E. Loginov

AN ARCHITECTURE OF AN ENTERPRISE MOBILITY MANAGEMENT SUITE FOR EDUCATION ESTABLISHMENTS

Интеграция мобильных устройств и сервисов в корпоративную инфраструктуру – актуальная для многих организаций задача, для решения которой чаще всего используются системы управления мобильностью предприятия (ЕММ). При внедрении системы ЕММ в образовательных учреждениях необходимо учитывать ряд специфических требований, связанных как с особенностями бизнес-процессов, так и с существенными отличиями между разными образовательными учреждениями. В статье проанализированы эти требования и предложена разработанная на их основе архитектура системы ЕММ. Особенностью этой архитектуры является возможность гибкого конфигурирования системы с учетом структуры и потребностей организации.

МОБИЛЬНЫЕ УСТРОЙСТВА; УПРАВЛЕНИЕ МОБИЛЬНЫМИ УСТРОЙСТВАМИ; АРХИТЕКТУРА; ОБРАЗОВАТЕЛЬНЫЕ УЧРЕЖДЕНИЯ.

Integration of mobile devices and services into the corporate IT infrastructure is an essential task for many organizations. The main approach to this problem involves an enterprise mobility management (EMM) system. Commercially available EMM systems share a relatively common set of basic features that are necessary for almost any organizations. Nevertheless, a specific business may require an EMM system to implement custom functionality to support its processes. In this paper, we consider the requirements specific to the domain of education. The most important requirement is the configurability of the EMM system, as schools and universities differs greatly in the amount of students and staff, and in the structure of the educational process. We propose an EMM system architecture that matches these requirements and allows constructing a scalable and configurable solution for mobility management in education.

MOBILE DEVICES; ENTERPRISE MOBILITY MANAGEMENT; SYSTEM ARCHITECTURE; EDUCATION.

Наличие мобильных устройств у большинства сотрудников стимулирует организации активно применять их в бизнес-процессах. Для этого используются специализированные решения, позволяющие безопасно и эффективно интегрировать личные мобильные устройства в корпоративную ИТ-среду. Интеграция мобильных устройств и технологий социаль-

ных сетей в корпоративную среду сегодня является главным стратегическим направлением развития в области информационных технологий [1].

Наиболее популярный подход к решению задачи интеграции – применение систем управления мобильностью предприятия (Enterprise Mobility Management – ЕММ). В соответствии с популярным опре-

делением компании Gartner, ЕММ — это класс программных средств, обеспечивающих использование мобильных устройств в корпоративных деловых процессах путем интеграции этих аппаратных средств в ИТ-системы и в среды обеспечения безопасности на всех этапах управления жизненным циклом ИТ [2].

Средства управления мобильностью предприятия ориентированы на решение комплекса задач по управлению мобильными устройствами (Mobile Device Management — MDM), мобильными приложениями (Mobile Application Management — MAM) и контентом (Mobile Content Management — MCM). На основании исследования большого количества средств этого класса Gartner выделяет следующий набор основных функций ЕММ [3].

- Управление мобильными устройствами: подготовка устройств к работе (удаленное конфигурирование мобильных устройств, операционных систем, настройка Wi-Fi, VPN, APN), управление инвентаризацией мобильных устройств, контроль соответствия мобильных устройств корпоративным политикам, геолокация.

- Управление мобильными приложениями: установка и удаление, управление конфигурацией и политиками приложений, контроль соответствия приложений корпоративным политикам, учет использования приложений, защита приложений, «черные списки» приложений.

- Управление контентом: обеспечение доступа к данным, защита корпоративных данных, управление правами доступа, шифрование данных, контроль информационных потоков, предоставление защищенного доступа к внутренним ИТ-ресурсам, отслеживание угроз безопасности данных (потеря устройств, неавторизованное реконфигурирование, увольнение сотрудников), удаленное устранение проблем, удаленное стирание (частичная или полная очистка данных при компрометации устройства).

Одним из главных факторов, влияющих на устройство систем ЕММ, является необходимость поддержки большого количества одновременно представленных на рынке мобильных систем разных производителей,

реализующих различные аппаратные и программные решения. Кроме того, системы ЕММ должны удовлетворять не только базовым, но и специфическим требованиям, которые определяются конкретными областями их применения [4, 5]. Указанные обстоятельства приводят к различным архитектурным решениям, применяемым при разработке ЕММ-инструментов.

Эта статья посвящена разработке системы управления мобильными устройствами, ориентированной на использование в образовательных учреждениях. Область применения выдвигает существенные требования, связанные как с поддерживаемой функциональностью, так и с нефункциональными характеристиками. В статье описываются технические решения, принятые при разработке системы. Они позволяют системе удовлетворять как общим требованиям к системам ЕММ, так и более специфическим требованиям, накладываемым областью применения. Насколько нам известно, ранее архитектуры систем управления мобильными устройствами, удовлетворяющих таким требованиям, в литературе не описывались.

Функциональность

Основная область применения разрабатываемой системы ЕММ — образовательные учреждения. Они имеют ряд особенностей, которые необходимо учитывать при проектировании системы [6].

- Наличие большого количества пользователей, каждый из которых нуждается в индивидуальном наборе приложений и доступе к разнообразному контенту. Это условие приводит к требованию расширенной функциональности уровней МAM и МСМ. В то же время отсутствие специальных требований к безопасности позволяет ограничиться базовыми функциями уровня MDM.

- Существование организаций разного размера: от нескольких сотен обучающихся и сотрудников в школах и колледжах до десятков тысяч студентов и сотрудников в больших университетах. Кроме величины, эти организации обычно обладают разными требованиями к функциональности,

обусловленными различиями реализуемых образовательных программ и контингента обучающихся. Все это приводит к необходимости разработки масштабируемых решений, обладающих развитыми возможностями настройки на конкретную образовательную организацию.

- Необходимость одновременного выполнения отдельных приложений большим количеством пользователей. Например, сервис контроля посещаемости в начале занятий может использоваться большинством студентов университета.

- Необходимость развитой системы групповых политик (например, в рамках каждой дисциплины студенты могут иметь доступ к различным ресурсам).

- Необходимость интеграции с различными информационными системами образовательного учреждения.

- Потребность в развитых средствах коммуникации между пользователями: короткие сообщения, электронная почта, корпоративные социальные сети.

- Необходимость развитых средств управления корпоративными приложениями в условиях многообразия аппаратных и

программных платформ: в частности, наличие корпоративных магазинов приложений для разных мобильных операционных систем.

- Наличие нескольких категорий пользователей (учащиеся, преподаватели, административный и вспомогательный персонал), для которых характерны различные сценарии использования устройств, а также использование устройств различных типов (смартфоны, планшеты, ноутбуки, стационарные компьютеры).

Архитектура системы управления мобильными устройствами

Результаты исследования 20 наиболее популярных средств ЕММ [5] показывают, что базовая функциональность поддерживается большей частью систем (табл. 1).

Архитектура существующих средств ЕММ [6–9], реализующая базовую функциональность, включает в себя программы-агенты, выполняемые на мобильных устройствах, управляющий сервер ЕММ, сервер базы данных, портал самообслуживания, почтовую службу. В качестве службы уведомлений системы ЕММ

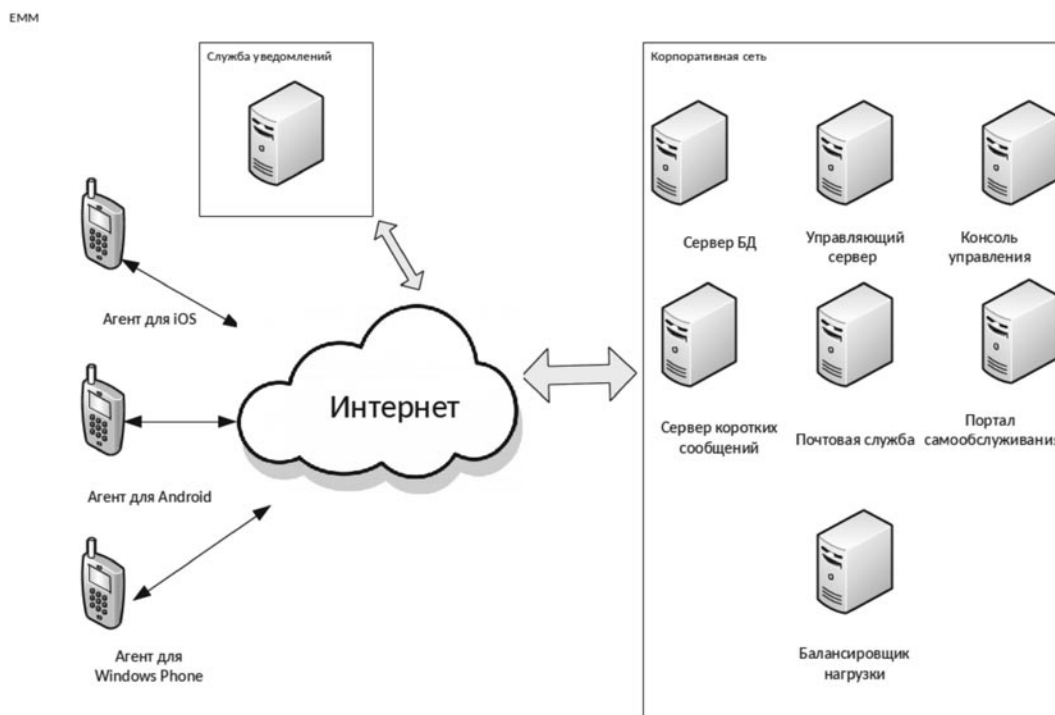


Рис. 1. Схема системы управления мобильными устройствами для образовательных организаций

Таблица 1

Поддержка основных функций известными реализациями ЕММ

Основная функциональность		Доля поддерживающих систем, %
Поддержка основных мобильных платформ	Windows Phone	90
	Android	100
	iOS	100
Используемые способы поставки	Облачное решение	80
	Локальное решение	90
Управление мобильными устройствами	Удаленное конфигурирование	100
	Управление инвентаризацией мобильных устройств	90
	Контроль соответствия мобильных устройств корпоративным политикам	80
	Геолокация и применение политик по местоположению	40
	Блокировка устройств	100
	Блокировка сенсоров	90
Управление мобильными приложениями	Инициализация и деинициализация приложений	100
	Управление конфигурацией и политиками приложений	90
	Учет использования приложений	80
Управление контентом	Обеспечение защищенного доступа к данным; контроль информационных потоков	100
	Предоставление защищенного доступа к внутренним ИТ-ресурсам	80
	Отслеживание угроз для данных	80

Таблица 2

Реализация в известных ЕММ возможностей, характерных для сферы образования

Основная функциональность	Доля поддерживающих функциональность средств, %
Поддержка большого числа (до 10 000) пользователей	55
Наличие большого количества (до 100) типов пользователей	60
Развитая система групповых политик	45
Корпоративные магазины приложений для разных платформ	40
Наличие большого количества средств коммуникации между пользователями	65
Интеграция с различными информационными системами	75

используют соответствующие службы поддерживаемых платформ (Apple Push Notification service, Google Cloud Messaging, Microsoft Push Notification Service).

Функциональность, определяемая требованиями предметной области, поддерживается существенно реже [3, 5, 10, 11]. Табл. 2 содержит обобщенные сведения о реализации соответствующих возможностей для наиболее популярных средств ЕММ.

Следует отметить, что всем требованиям предметной области не удовлетворяет ни одно из исследованных средств.

Необходимость удовлетворения перечисленных выше требований, которые определяются областью применения системы, привела к ряду архитектурных решений. В качестве основного принципа организации был выбран сервисно-ориентированный подход. Отдельные элементы функциональности реализуются в виде модулей, связываемых между собой посредством протоколов взаимодействия. Это дает возможность формирования нужного набора функций как путем удаления неиспользуемых модулей, так и путем добавления новых модулей, реализующих ранее не предусмотренные возможности. Сформированный набор модулей покрывает всю заявленную функциональность и позволяет легко расширять ее, если возникнет необходимость. Такое решение обеспечивает адаптацию разрабатываемых средств к требованиям различных образовательных учреждений.

Необходимость поддерживать разные типы клиентских устройств требует поддержки разных типов клиентских приложений (собственные приложения мобильных платформ, веб-клиенты, приложения для настольных операционных систем). Такая возможность достигается за счет открытого унифицированного API и интерфейса REST.

Унифицированная поддержка различных мобильных платформ реализуется с помощью введения для каждой платформы специального встраиваемого модуля для подключения мобильного устройства, трансляции команд сервера ЕММ в команды управления мобильным устройством и в обратном направлении, а также для взаи-

модействия с модулем управления мобильными устройствами, входящими в состав мобильной ОС.

С учетом перечисленных архитектурных решений система управления мобильными устройствами для образовательных учреждений состоит из следующих компонентов (см. рис. 1).

1. *Сервер базы данных.* Обеспечивает хранение информации о пользователях и мобильных устройствах, хранение инвентаризационной информации, конфигурационных данных, политик, распространяемых приложений и документов. Взаимодействие компонентов системы с сервером базы данных осуществляется через управляющий сервер.

2. *Консоль управления.* Необходимость централизованного управления мобильными устройствами пользователей со стороны ИТ-службы образовательного учреждения определила включение в архитектуру системы компонента «консоль управления», который служит для включения, выключения и настройки всех функций системы ЕММ. Консоль управления предоставляет графический интерфейс администратора для всех модулей управляющего сервера.

3. *Сервер коротких сообщений.* Необходимость массовой рассылки сообщений (например, в случае экстренных оповещений) реализуется с помощью специального компонента «Сервер рассылки коротких сообщений», работающего независимо от основного сервера ЕММ.

4. *Почтовая служба.* Стандартный сервис, обеспечивающий функционирование корпоративной электронной почты.

5. *Портал самообслуживания.* Обеспечивает доступ пользователей к ряду функций ЕММ. При помощи портала самообслуживания пользователь может самостоятельно зарегистрировать мобильные устройства и отправлять на свои мобильные устройства некоторые команды, такие как блокировка экрана, сброс кода разблокировки, временное отключение устройства от системы, просмотр местоположения, проигрывание звукового сигнала. Помимо этого, через портал самообслуживания пользователь имеет доступ к функциям отправки и полу-

чения сообщений, просмотра корпоративных документов, просмотра списка доступных веб-сервисов.

6. *Сервер балансировки нагрузки.* Для обеспечения одновременной работы большого количества пользователей в состав системы введены средства балансировки нагрузки. Сервер балансировки нагрузки осуществляет распределение нагрузки между различными развернутыми копиями управляющего сервера, перенаправляя запросы мобильных устройств на экземпляр управляющего сервера, наилучшим образом удовлетворяющий установленным администратором системы критериям. Сервер балансировки нагрузки также осуществляет мониторинг состояния развернутых экземпляров управляющих серверов, чтобы в случае выхода из строя отдельного экземпляра можно было исключить его из процесса обработки запросов мобильных устройств.

7. *Приложение «Агент» для различных мобильных платформ.* Предоставляет конечному пользователю следующие функции [12]: получение и отправка сообщений другим пользователям системы, просмотр каталога доступных документов, загрузка и обновление доступных документов, загрузка, установка и обновление доступных корпоративных приложений, просмотр информации о мероприятиях, участником которых является пользователь, регистрация на мероприятиях. В зависимости от мобильной платформы, перечисленный набор функций может меняться. Например, в приложении «Агент» для платформы Android также встроены функции MDM-агента, т. е. компонента, отвечающего за применение заданных на управляющем сервере политик и выполнение полученных от сервера команд.

Основным компонентом системы является управляющий сервер. Он включает в себя следующие модули (рис. 2):

- управления пользователями и группами;
- управления политиками для мобильных устройств;
- управления приложениями для мобильных устройств;
- управления сообщениями;

- управления документами;
- управления мероприятиями;
- рассылки уведомлений.

Взаимодействие между модулями происходит в соответствии с разработанными протоколами взаимодействия и межмодульными интерфейсами. Каждый отдельный модуль позволяет обеспечивать выполнение определенного класса задач или бизнес-процессов. Например, модуль управления сообщениями, входящий в состав управляющего сервера, позволяет пользователям системы отправлять сообщения другим пользователям, не зная их контактных данных (номера телефона, адреса электронной почты). Модуль управления документами позволяет пользователям получать новую рабочую документацию сразу, как только она стала доступна, не требуя от пользователя отдельно забирать ее из файлового хранилища компании.

Модуль управления пользователями и группами. Модуль позволяет регистрировать в системе новых пользователей и редактировать параметры существующих пользователей. Также модуль реализует функции импорта пользователей из других информационных систем с помощью стандартных интерфейсов (например, LDAP). Администратор системы может путем вызова соответствующей функции данного модуля блокировать и разблокировать отдельных пользователей системы. Помимо управления пользователями, модуль позволяет создавать новые группы пользователей, редактировать существующие группы, а также управлять членством пользователей в группах и формировать иерархии групп.

Модуль управления политиками для мобильных устройств. Задачей этого модуля является контроль прав доступа, назначенных отдельным пользователям и группам пользователей. Модуль позволяет задавать политики для мобильных устройств, включающие следующие виды ограничений:

- запрет или разрешение использования камеры мобильного устройства;
- формирование требований к сложности пароля для разблокировки мобильного устройства;
- возможность установки приложений;

запрет или разрешение снятия снимков экрана;

запрет или разрешение синхронизации в роуминге;

необходимость шифрования данных;
автоматизированная настройка электронной почты;

автоматическая настройка удаленного доступа к корпоративной сети (VPN);

задание перечня доступных веб-сервисов.

Модуль управления приложениями для мобильных устройств. Модуль обеспечивает управление корпоративными приложениями для мобильных устройств и веб-приложениями. Он позволяет автоматически развертывать приложения, необходимые для выполнения конкретным пользователем его должностных обязанностей или учебного плана, а также рекомендо-

вать установку тех или иных приложений на основе членства пользователя в группах и сообщать о наличии новых версий таких приложений.

Модуль управления мероприятиями. Модуль несколько отличается от других тем, что предназначен для решения задачи более высокого уровня. Тем не менее управление мероприятиями реализовано как модуль системы, а не как отдельный сервис, поскольку мероприятия (учебные занятия, семинары, конференции) являются основной технологической единицей образовательного процесса, и поэтому функции управления мероприятиями необходимы для работы других компонентов системы. В частности, со средствами управления мероприятиями интегрированы средства управления документами, обеспечивающие индивидуальный и групповой доступ

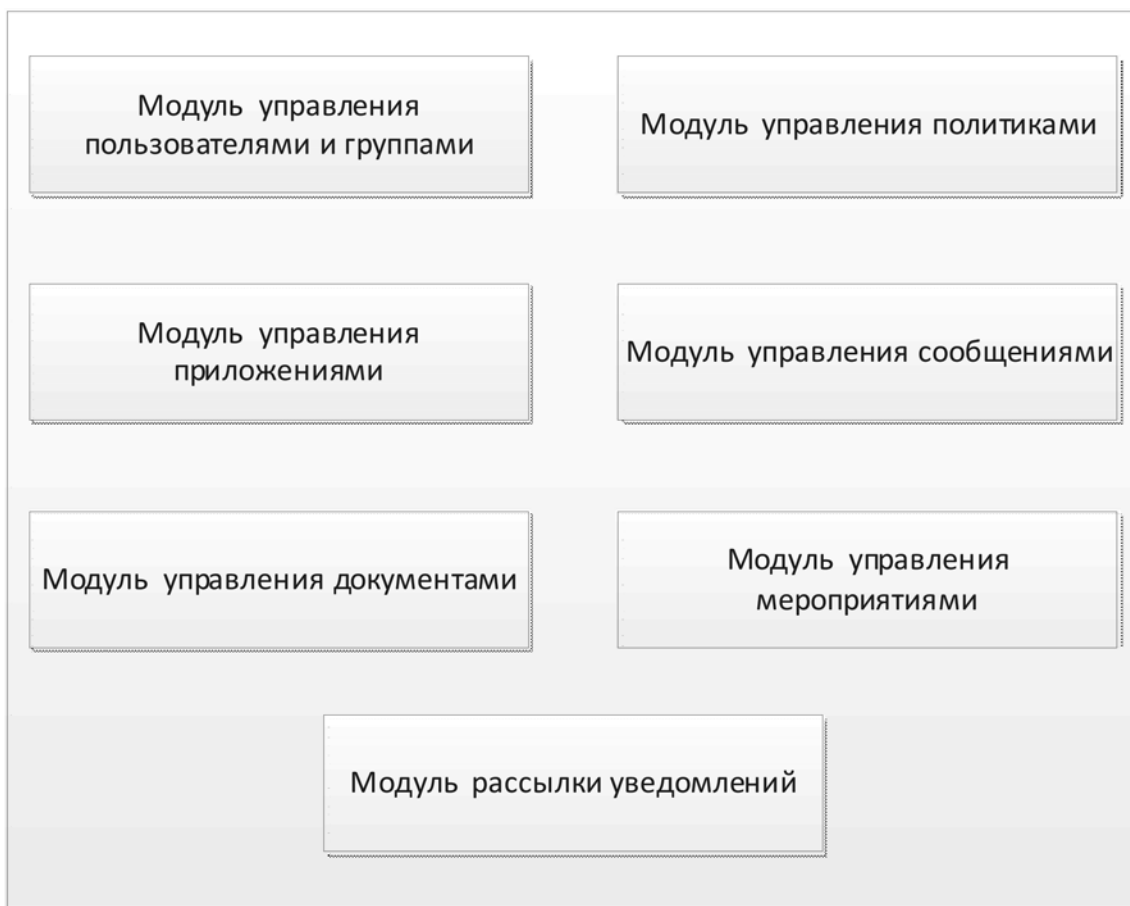


Рис. 2. Модульная структура управляющего сервера

пользователей к учебным материалам и каталогам документов с учетом учебных планов и выполняемых сотрудниками обязанностей. Модуль позволяет создавать обычные и повторяющиеся мероприятия, назначать ведущих и участников мероприятий и задавать политики, которые будут действовать во время проведения мероприятий.

Модуль управления сообщениями. Модуль позволяет отправлять и получать сообщения, обеспечивая индивидуальную и групповую коммуникацию между пользователями системы. С помощью данного модуля можно управлять правами пользователей на отправку сообщений группам.

Модуль управления документами. Модуль позволяет загружать в систему электронные документы, управлять правами доступа к документам, автоматически рассылать рекомендуемые документы на устройства пользователей, а также информировать пользователей об изменениях в документах.

Модуль рассылки уведомлений. Модуль осуществляет рассылку уведомлений на мобильные устройства пользователей системы. Он агрегирует уведомления от других компонентов сервера и осуществляет централизованную отправку системных уведомлений через общедоступные каналы связи.

Взаимодействие управляющего сервера с другими компонентами осуществляется с помощью программного интерфейса Web-API. Реализовано два таких интерфейса: один предназначен для мобильных платформ, другой – для консоли управления и портала самообслуживания.

Для каждой поддерживаемой мобильной платформы в системе существует свой набор доступных интерфейсов, через которые осуществляется управление мобильными устройствами. Всё взаимодействие мобильных устройств с управляющим сервером происходит только посредством этих интерфейсов. Для консоли управления и портала самообслуживания предусмотрены дополнительные интерфейсы, предоставляющие доступ к служебным методам сервера.

Результаты работы

Практическим результатом является разработка программной системы «СмартГейт», реализующей описанную архитектуру. Все функциональные возможности платформы проверены с помощью разработанных инструментальных средств поддержки тестирования для систем централизованного управления мобильными устройствами [13]. Нагрузочные тесты показали, что платформа «СмартГейт» обеспечивает корректное функционирование при нагрузке, что позволяет, с учетом распределения обращений к серверу во времени, поддерживать работу порядка 30 000 пользователей.

Апробация системы, в которой участвовали шесть преподавателей и 128 студентов, была проведена в Санкт-Петербургском политехническом университете Петра Великого (СПбПУ). При этом в учебном процессе были использованы возможности системы по регистрации пользователей, регистрации мобильных устройств, созданию групп пользователей и групповых политик, распространению учебных материалов, приложений для учебного процесса, различным видам коммуникаций. Кроме этого, на базе созданной платформы были созданы и проверены сервис контроля посещаемости и сервис проведения тестов и опросов. Апробация проводилась в реальных условиях учебного процесса и позволила убедиться в корректности поведения платформы и сервисов при работе с мобильными устройствами разных моделей с персональными настройками пользователей под управлением всех основных операционных систем. Таким образом, результаты тестирования и апробации свидетельствуют о соответствии принятых архитектурных решений требованиям.

В статье предложена архитектура системы управления мобильными устройствами для образовательных организаций, включающей в себя совокупность сервисов, интерфейсы взаимодействия и клиентские приложения. Система построена по модульному принципу и обеспечивает кон-

фигурирование с учетом конкретных функциональных требований.

Реализованная на основе данной архитектуры платформа «СмартГейт» обеспечивает всю базовую функциональность систем ЕММ, а также предоставляет дополнительные возможности, которые необходимы для решения специфических задач, возникающих при внедрении мобильных сервисов и технологий в образовательных учреждениях.

Нагрузочное тестирование системы «СмартГейт» и ее апробация в условиях реального учебного процесса СПбПУ продемонстрировали масштабируемость системы

при росте нагрузки и корректную поддержку различных мобильных платформ.

Разработка программного обеспечения системы централизованного управления мобильными устройствами, лежащей в основе комплекса, проводится в рамках совместного проекта компании IBS (Москва) и Санкт-Петербургского политехнического университета Петра Великого.

Работа подготовлена в ходе реализации комплексного проекта в рамках Постановления Правительства РФ от 09.04.2010 г. № 218 при финансовой поддержке Министерства образования и науки РФ. Договор № 02.G25.31.0024 от 12.02.2013 г.

СПИСОК ЛИТЕРАТУРЫ

1. **Агапов А., Пратусевич В., Яковлев С.** Обзор и оценка перспектив развития мирового и российского рынков информационных технологий. IDC, 2014.
2. **Taylor B., Basso M., Wong J.** IT Market Clock for Enterprise Mobility. Gartner, 2014 [электронный ресурс]/ URL: <https://www.gartner.com/doc/2854320?ref=SiteSearch&sthkw=EMM&fnl=search&srcId=1-3478922254> (дата обращения: 15.10.2015).
3. **Cosgrove T., Smith R., Silva C., Girard J., Taylor B.** Magic Quadrant for Enterprise Mobility Management Suites. Gartner, 2015.
4. **Majdi E.B.** Evaluation of mobile device management tools and analyzing integration models for mobility enterprise: dissertation. Umee, Sweden: Umee university, 2013.
5. **Самочадин А.В., Самочадина Т.Н., Тимофеев Д.А., Попов С.О.** Базовая функциональность систем централизованного управления мобильными устройствами // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. СПб.: Изд-во Политехн. ун-та, 2014. № 4 (200). С. 17–28.
6. **Samochadin A., Raychuk D., Voinov N., Ivanchenko D., Khmelkov I.** MDM-based Mobile Services in Universities // Internat. J. of Information Technology & Computer Science. 2014. Vol. 13. No. 2. Pp. 35–41.
7. McAfee Enterprise Mobility Management [электронный ресурс]/ URL: <http://www.itprotect.ru/product/mcafee/mcafee-enterprise-mobility-management/> (дата обращения: 14.10.2015).
8. Mobile Device Architecture [электронный ресурс]/ URL: <https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Tivoli+Endpoint+Manager/page/Mobile+Device+Architecture> (дата обращения: 14.10.2015).
9. Mobile Device Manager. Device Management Server Architecture [электронный ресурс]/ URL: <https://technet.microsoft.com/en-us/library/dd261915.aspx> (дата обращения: 14.10.2015).
10. Check Point. The Myths of mobile Security [электронный ресурс]/ URL: <https://bacher.at/assets/Newsletter/2015-03-CheckPoint-TestReport.pdf> (дата обращения: 16.11.2015).
11. Enterprise Mobility Management (EMM) Security. 451 research [электронный ресурс]/ URL: <https://www.air-watch.com/downloads/resources/voice-of-the-enterprise-information-security-slides-q2-2015-airwatch-by-vmware.pdf> (дата обращения: 16.11.2015).
12. **Райчук Д.Ю., Самочадин А.В., Носницын С.А., Хмельков И.А.** Комплекс мобильных средств поддержки учебного процесса // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. СПб.: Изд-во Политехн. ун-та, 2014. № 6 (210). С. 7–15.
13. **Самочадин А.В., Сужаев О.И., Тимофеев Д.А., Рогов П.А.** Инструментальные средства нагрузочного тестирования для систем централизованного управления мобильными устройствами // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. СПб.: Изд-во Политехн. ун-та, 2014. № 6 (210). С. 33–41.

REFERENCES

1. **Agapov A., Pratushevich V., Yakovlev S.** *Obzor i otsenka perspektiv razvitiya mirovogo i rossiyskogo rynkov informatsionnykh tekhnologiy* [Review and assessment of the prospects of development of the world and Russian markets of Information Technology], IDC, 2014. (rus)
2. **Taylor B., Basso M., Wong J.** *IT Market Clock for Enterprise Mobility*. Gartner, 2014. Available: <https://www.gartner.com/doc/2854320?ref=SiteSearch&stkw=EMM&fnl=search&srcId=1-3478922254> (Accessed: 15.10.2015).
3. **Cosgrove T., Smith R., Silva C., Girard J., Taylor B.** *Magic Quadrant for Enterprise Mobility Management Suites*. Gartner, 2015.
4. **Majdi E.B.** *Evaluation of mobile device management tools and analyzing integration models for mobility enterprise: dissertation*. Umee, Sweden: Umee university, 2013.
5. **Samochadin A.V., Samochadina T.N., Timofeev D.A., Popov S.O.** Bazovaya funktsionalnost sistem tsentralizovannogo upravleniya mobilnymi ustroystvami [Base Features of Mobile Device Management Systems]. *Nauchno-tehnicheskie vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravlenie* [St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control System], St. Petersburg: SPbGPU Publ., 2014, No. 4 (200), Pp. 17–28. (rus)
6. **Samochadin A., Raychuk D., Voinov N., Ivanchenko D., Khmelkov I.** MDM-based Mobile Services in Universities. *Internat. J. of Information Technology & Computer Science*, 2014, Vol. 13, No. 2, Pp. 35–41.
7. **McAfee Enterprise Mobility Management**. Available: <http://www.itprotect.ru/product/mcafee/mcafee-enterprise-mobility-management/> (Accessed: 14.10.2015).
8. **Mobile Device Architecture**. Available: <https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Tivoli+Endpoint+Manager/page/Mobile+Device+Architecture> (Accessed: 14.10.2015).
9. **Mobile Device Manager. Device Management Server Architecture**. Available: <https://technet.microsoft.com/en-us/library/dd261915.aspx> (Accessed: 14.10.2015).
10. **Check Point. The Myths of mobile Security**. Available: <https://bacher.at/assets/Newsletter/2015-03-CheckPoint-TestReport.pdf> (Accessed: 16.11.2015).
11. **Enterprise Mobility Management (EMM) Security. 451 research**. Available: <https://www.air-watch.com/downloads/resources/voice-of-the-enterprise-information-security-slides-q2-2015-airwatch-by-vmware.pdf> (Accessed: 16.11.2015).
12. **Raychuk D.Yu., Samochadin A.V., Nosnitsyn S.M., Khmelkov I.A.** Kompleks mobilnykh sredstv podderzhki uchebnogo protsessa [A Set of Mobile Applications to Support the Education Process]. *Nauchno-tehnicheskie vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravlenie* [St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control System], St. Petersburg: SPbGPU Publ., 2014, No. 6 (210), Pp. 7–15. (rus)
13. **Samochadin A.V., Suzhaev O.I., Timofeev D.A., Rogov P.A.** Instrumentalnyiye sredstva nagruzochnogo testirovaniya dlya sistem tsentralizovannogo upravleniya mobilnymi ustroystvami [Tools for load testing Mobile Device Management systems]. *Nauchno-tehnicheskie vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravlenie* [St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control System], St. Petersburg: SPbGPU Publ., 2014, No. 6 (210), Pp. 33–41. (rus)

МАСЛОВ Максим Юрьевич – заместитель заведующего лабораторией систем управления мобильными устройствами Санкт-Петербургского политехнического университета Петра Великого.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: maslov@soft-consult.ru

MASLOV Maxim Yu. Peter the Great St. Petersburg Polytechnic University.

195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.

E-mail: maslov@soft-consult.ru

НОСНИЦЫН Семен Михайлович – аспирант кафедры распределенных вычислений и компьютерных сетей Института информационных технологий и управления Санкт-Петербургского политехнического университета Петра Великого.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: lgalod@gmail.com

NOSNITSYN Semen M. *Peter the Great St. Petersburg Polytechnic University.*
195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.
E-mail: lgalod@gmail.com

САМОЧАДИН Александр Викторович – профессор кафедры распределенных вычислений и компьютерных сетей Института информационных технологий и управления Санкт-Петербургского политехнического университета Петра Великого, кандидат технических наук.
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: samochadin@gmail.com

SAMOCHADIN Alexander V. *Peter the Great St. Petersburg Polytechnic University.*
195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.
E-mail: samochadin@gmail.com

ЛОГИНОВ Кирилл Евгеньевич – руководитель проектов, *IBS.*
127434, Россия, Москва, Дмитровское ш., д. 9Б.
E-mail: KLoginov@ibs.ru

LOGINOV Kirill E. *IBS.*
127434, Dmitrovskoe shosse, 9b, Moscow, Russia.
E-mail: KLoginov@ibs.ru

НАУЧНОЕ ИЗДАНИЕ
«НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ
САНКТ-ПЕТЕРБУРГСКОГО
ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА.
ИНФОРМАТИКА. ТЕЛЕКОММУНИКАЦИИ. УПРАВЛЕНИЕ»
«ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL.
COMPUTER SCIENCE. TELECOMMUNICATIONS AND CONTROL SYSTEMS»

№ 5 (229) 2015

Учредитель – Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Санкт-Петербургский государственный политехнический университет»

Журнал зарегистрирован Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор).
Свидетельство о регистрации ПИ № ФС77-51457 от 19.10.2012 г.

Редакция журнала

д-р техн. наук, профессор *А.С. Коротков* – главный редактор
Е.А. Калинина – литературный редактор, корректор
Г.А. Пышкина – ответственный секретарь, выпускающий редактор

Телефон редакции (812)552-62-16, 297-18-21

E-mail: infocom@spbstu.ru

Компьютерная верстка *А.Н. Смирнов*

Директор Издательства Политехнического университета *А.В. Иванов*

Лицензия ЛР № 020593 от 07.08.97

Подписано в печать 30.10.2015. Формат 60×84 1/8. Бум. тип. № 1.
Печать офсетная. Усл. печ. л. 13,02. Уч.-изд. л. 13,02. Тираж 1000. Заказ

Санкт-Петербургский политехнический университет Петра Великого
Издательство Политехнического университета
член Издательско-полиграфической ассоциации университетов России
Адрес университета и издательства: 195251, Санкт-Петербург, ул. Политехническая, д. 29.

УСЛОВИЯ ПУБЛИКАЦИИ СТАТЕЙ

в журнале «Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление»

1. ОБЩИЕ ПОЛОЖЕНИЯ

Журнал «Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Телекоммуникации. Управление» является периодическим печатным научным рецензируемым изданием. Зарегистрировано Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации ПИ № ФС77-51457 от 19 октября 2012 г. С 2008 года выпускается в составе сериального периодического издания «Научно-технические ведомости СПбГПУ» (ISSN 1994-2354).

Издание с 2002 года входит в Перечень ведущих научных рецензируемых журналов и изданий (перечень ВАК) и принимает для печати материалы научных исследований, а также статьи для опубликования основных результатов диссертаций на соискание ученой степени доктора наук и кандидата наук по следующим основным научным направлениям: **ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА, РАДИОТЕХНИКА И СВЯЗЬ, ЭЛЕКТРОНИКА, ИЗМЕРИТЕЛЬНАЯ ТЕХНИКА, УПРАВЛЕНИЕ В СОЦИАЛЬНЫХ И ЭКОНОМИЧЕСКИХ СИСТЕМАХ**. Научные направления журнала учитываются ВАК Минобрнауки РФ при защите докторских и кандидатских диссертаций в соответствии с Номенклатурой специальностей научных работников.

Сведения о публикациях представлены в РИНЦ, в Реферативном журнале ВИНИТИ РАН, в международной справочной системе «Ulrich's Periodical Directory».

Периодичность выхода журнала – 6 номеров в год.

2. ТРЕБОВАНИЯ К ПРЕДОСТАВЛЯЕМЫМ МАТЕРИАЛАМ

2.1. Оформление материалов

1. Рекомендуемый объем статей для авторов с ученой степенью доктора наук, званием профессора, соискателей ученой степени доктора наук (докторантов) 12–20 страниц формата А-4 с учетом графических вложений. Количество графических вложений (диаграмм, графиков, рисунков, таблиц, фотографий и т. п.) не должно превышать 4.

2. Рекомендуемый объем статей для преподавателей, авторов без ученой степени, соискателей ученой степени кандидата наук – 8–15 страниц формата А-4; аспирантов – 8 страниц формата А-4 с учетом графических вложений. Количество графических вложений (диаграмм, графиков, рисунков, таблиц, фотографий и т. п.) не должно превышать 3.

3. Авторы должны придерживаться следующей обобщенной структуры статьи: вводная часть (0,5–1 стр., актуальность, существующие проблемы); основная часть (постановка и описание задачи, изложение и суть основных результатов); заключительная часть (0,5–1 стр., предложения, выводы), список литературы (оформление по ГОСТ 7.05.-2008).

4. Число авторов статьи не должно превышать трех человек.

5. Набор текста осуществляется в редакторе **MS Word**, формул – в редакторе **MathType**. Таблицы набираются в том же формате, что и основной текст.

6. Шрифт – **TNR**, размер шрифта основного текста – 14, интервал – 1,5; таблицы большого размера могут быть набраны 12 кеглем. Параметры страницы: поля слева – 3 см, сверху, снизу – 2,5 см, справа – 2 см, текст размещается без переносов. Абзацный отступ – 1 см.

2.2. Предоставление материалов

Вместе с материалами статьи должны быть обязательно предоставлены:

- номер УДК в соответствии с классификатором (в заголовке статьи);
- аннотация на русском и английском языках;
- ключевые слова (5–7) на русском и английском языках;
- сведения об авторах на русском и английском языках: ФИО, место работы, должность, ученое звание, ученая степень, контактные телефоны, e-mail;
- аспиранты представляют документ отдела аспирантуры, заверенный печатью;
- акт экспертизы о возможности опубликования материалов в открытой печати.

С авторами статей заключается издательский лицензионный договор.

Предоставление всех материалов осуществляется в электронном виде через личный кабинет **ЭЛЕКТРОННОЙ РЕДАКЦИИ** по адресу <http://journals.spbstu.ru>

2.3. Рассмотрение материалов

Предоставленные материалы (п. 2.2) первоначально рассматриваются редакционной коллегией и передаются для рецензирования. После одобрения материалов, согласования различных вопросов с автором (при необходимости) редакционная коллегия сообщает автору решение об опубликовании статьи. В случае отказа в публикации статьи редакция направляет автору мотивированный отказ.

При отклонении материалов из-за нарушения сроков подачи, требований по оформлению или как не отвечающих тематике журнала материалы не публикуются и не возвращаются.

Редакционная коллегия не вступает в дискуссию с авторами отклоненных материалов.

Публикация материалов аспирантов очной бюджетной формы обучения осуществляется бесплатно в соответствии с очередностью.

При поступлении в редакцию значительного количества статей их прием в очередной номер может закончиться **ДОСРОЧНО**.

Более подробную информацию можно получить:

на сайте журнала <http://ntv.spbstu.ru>

по телефону редакции +7(812) 552-62-16 с 10⁰⁰ до 18⁰⁰ Галина Александровна

или по e-mail: infocom@spbstu.ru