



НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ

САНКТ-ПЕТЕРБУРГСКОГО ГОСУДАРСТВЕННОГО
ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

Информатика. Телекоммуникации.
Управление

Том 11, № 2
2018

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:

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;

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.

Prof. Dr. *Fa-Long Luo*, Affiliate Full Professor University of Washington, USA, Chief Scientist Micron Technology, Inc., Milpitas, USA, Chairman IEEE SPS Industry DSP Technology Standing Committee;

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;

Прокопенко Н.Н., д-р техн. наук, профессор, Донской государственный технический университет, г. Ростов-на-Дону, Россия;

Путря М.Г., д-р техн. наук, профессор, Национальный исследовательский университет «Московский институт электронной техники», Москва, Россия;

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

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

Уткин Л.В., 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;

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

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 ВИНИТИ РАН, в международной справочной системе «Ulrich's Periodical Directory», в базах данных Российский индекс научного цитирования (РИНЦ), Google Scholar, EBSCO, Math-Net.Ru, ProQuest, Index Copernicus

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.

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.



**ST. PETERSBURG STATE
POLYTECHNICAL UNIVERSITY
JOURNAL**

Computer Science.
Telecommunications and Control Systems

**Vol. 11, No. 2
2018**

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:

Assoc. Prof. Dr. *Vladimir M. Itsyson*, 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. *Yevgeni Koucheryavy*, Tampere University of Technology, Finland.

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

Prof. Dr. *Fa-Long Luo*, Affiliate Full Professor University of Washington, USA, Chief Scientist Micron Technology, Inc., Milpitas, USA, Chairman IEEE SPS Industry DSP Technology Standing Committee;

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. *Nikolay N. Prokopenko*, Don State Technical University, Rostov-on-Don, Russia;

Prof. Dr. *Mikhail G. Putrya*, National Research University of Electronic Technology, Moscow, Russia;

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

Assoc. Prof. Dr. *Peter V. Trifonov*, 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.

Prof. Dr. *Lev V. Utkin*, Peter the Great St. Petersburg Polytechnic University, Russia.

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 journal is indexed by Ulrich's Periodicals Directory, Google Scholar, EBSCO, ProQuest, Index Copernicus, VINITI RAS Abstract Journal (Referativnyi Zhurnal), VINITI RAS Scientific and Technical Literature Collection, Russian Science Citation Index (RSCI) database © Scientific Electronic Library and Math-Net.ru databases.

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.

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.

Содержание

Информационные технологии

Макаров А.С., Болсуновская М.В. Сравнительный анализ методов обнаружения особых точек на изображениях при различных уровнях освещения	7
--	---

Телекоммуникационные системы и компьютерные сети

Пяттаев А.В., Андреев С.Д., Кучерявый Е.А. Разработка и исследование системы прямых соединений D2D с сотовой поддержкой LTE	19
--	----

Программное обеспечение вычислительных, телекоммуникационных и управляющих систем

Моисеенко Е.А., Подкопаев А.В. Реляционное программирование с мемоизацией и отрицанием	35
---	----

Системный анализ и управление

Суляев И.И., Седов Д.В. Моделирование воздухоразделительной установки низкого давления как объекта управления	47
--	----

Интеллектуальные системы и технологии

Тянь Чжаолинь, Чжан Вэйвэй. Модель Sequence-to-Sequence в англо-китайском переводе	55
Сичкар В.Н. Сравнительный анализ систем, основанных на знаниях для навигации мобильного робота и предотвращения столкновений с препятствиями в неизвестной среде	64

Contents

Information Technologies

Makarov A.S., Bolsunovskaya M.V. <i>Comparative analysis of methods for detecting special points in images at different levels of illumination</i>	7
---	---

Telecommunication Systems and Computer Networks

Pyattaev A.V., Andreev S.D., Koucheryavy Ye.A. <i>Design and evaluation of a system of direct D2D connections with LTE cellular assistance</i>	19
---	----

Computer Software, Telecommunications, and Control Systems

Moiseenko E.A., Podkopaev A.V. <i>Relational programming with memoization and negation</i>	35
---	----

System Analysis and Control

Sulyaev I.I., Sedov D.V. <i>Modeling of a low-pressure air separator as a control object</i>	47
---	----

Intellectual Systems and Technologies

Tian Zhaolin, Zhang Weiwei. <i>Sequence-to-Sequence based english-chinese translation model</i>	55
Sichkar V.N. <i>Comparison analysis of knowledge-based systems for navigation of mobile robot and collision avoidance with obstacles in unknown environment</i>	64

DOI: 10.18721/JCSTCS.11201

УДК 004.932

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ОБНАРУЖЕНИЯ ОСОБЫХ ТОЧЕК НА ИЗОБРАЖЕНИЯХ ПРИ РАЗЛИЧНЫХ УРОВНЯХ ОСВЕЩЕНИЯ

А.С. Макаров, М.В. Болсуновская

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

Проведен сравнительный анализ методов нахождения особых точек на изображении, являющийся частью работы по созданию системы кругового обзора для крупногабаритных транспортных средств. Так как особенно опасным для вождения и наиболее сложным для проведения сшивки изображений является темное время, особое внимание в данной статье уделено возможности поиска особых точек и сшивки при низком освещении. Разработана методика и проведены эксперименты по нахождению особых точек на изображениях методами SURF, MSER, BRISK, Harris, FAST и MinEigen. Осуществлен поиск общих особых точек для пары изображений, а также проведены анализ их количества и сшивка изображений разными методами при различных уровнях освещения.

Ключевые слова: поиск особых точек, сшивка изображений, низкий уровень освещения, методика сравнения, SURF.

Ссылка при цитировании: Макаров А.С., Болсуновская М.В. Сравнительный анализ методов обнаружения особых точек на изображениях при различных уровнях освещения // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2018. Т. 11. № 2. С. 7–18. DOI: 10.18721/JCSTCS.11201

COMPARATIVE ANALYSIS OF METHODS FOR DETECTING SPECIAL POINTS IN IMAGES AT DIFFERENT LEVELS OF ILLUMINATION

A.S. Makarov, M.V. Bolsunovskaya

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

In this article, we have carried out a comparative analysis of methods for detecting special points in an image, which is part of the study on developing an around view system for large vehicles. Because nighttime is especially dangerous for driving and most difficult for stitching images, this article focused on the possibility of detecting

special points and stitching in low illumination. We made a comparative analysis of the methods for detecting special points in images, developed a technique and conducted experiments to find special points in images using such methods as SURF, MSER, BRISK, Harris, FAST and MinEigen. During the study, we have performed a search for identical special points for a pair of images, an analysis of their number and stitching of images by different methods at different levels of illumination. The results of the experiments are given in graphs and tables.

Keywords: detecting special points, stitching images, low level of illumination, comparison technique, SURF.

Citation: Makarov A.S., Bolsunovskaya M.V. Comparative analysis of methods for detecting special points in images at different levels of illumination. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems, 2018, Vol. 11, No. 2, Pp. 7–18. DOI: 10.18721/JCSTCS.11201

Введение

В последнее время стало уделяться много внимания автоматизации управления безопасностью на дорогах.

Вождение сопряжено с неконтролируемым риском малых столкновений и наездов на пешеходов из-за «слепых зон». Ежегодно на дорогах Европы в авариях, в которых участвуют грузовые автомобили, погибает более семи тысяч человек, а около ста тысяч получают повреждения [1]. Европейская Комиссия и представители Международного союза автомобильного транспорта (IRU) провели исследование, в ходе которого выяснилось, что около 75 % дорожно-транспортных происшествий с участием грузовых транспортных средств случаются именно из-за «слепых зон» [2].

Основной целью исследования является разработка системы кругового обзора для крупногабаритных транспортных средств. Результатом должен стать программно-аппаратный комплекс, состоящий из четырех широкоугольных fish-eye камер, аппаратного блока, монитора, интерфейса и программного обеспечения, реализующего разработанные алгоритмы обработки видеосигнала.

Данный комплекс, получая изображения с четырех камер, выведет на монитор в режиме реального времени единое изображение ТС и его окружения «с высоты птичьего полета».

Получение такого изображения состоит из двух этапов:

1. Получение ортогографического вида

сверху с четырех камер (этот этап описан в статьях [3, 4]).

2. Сшивка изображений, полученных на первом этапе, в единое изображение.

В настоящей статье рассматриваются подходы к реализации второго этапа. Вследствие того, что наиболее опасным для вождения и сложным для проведения сшивки изображений является темное время суток, особое внимание уделено возможности поиска особых точек и сшивки изображений при низком освещении.

Обзор существующих методов

В данном разделе приведено краткое описание и сравнительный анализ существующих методов поиска особых точек на изображении. Наиболее часто используемыми методами являются SURF и SIFT.

Метод SURF (Speeded up Robust Features) [5] обычно применяется для решения двух задач: поиска особых точек на изображении и создания их дескрипторов (описательного элемента, инвариантного к изменению масштаба и поворота). Кроме того, сам поиск ключевых точек тоже должен обладать инвариантностью, т. е. повернутый объект сцены должен обладать тем же набором ключевых точек, что и образец.

Метод ищет особые точки с помощью матрицы Гессе. Детерминант матрицы Гессе (т. н. гессиан) достигает экстремума в точках максимального изменения градиента яркости. Для двумерной функции ее детерминант определяется следующим образом [6]:

$$H(f(x, y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}, \quad (1)$$

$$\det(H) = \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2,$$

где H – матрица Гессе; $f(x, y)$ – функция изменения градиента яркости.

Стандартная версия SURF в несколько раз быстрее, чем SIFT, и, как утверждают ее авторы, более инвариантна против различных преобразований изображений, чем SIFT [5].

В *методе SIFT* (Scale-Invariant Feature Transform) [7] ключевые точки объектов сначала извлекаются из набора опорных изображений и сохраняются в базе данных. Объект распознается в новом изображении путем индивидуального сравнения каждого признака нового изображения с базой данных и поиска подходящих кандидатов на основе евклидова расстояния их векторов признаков. Из полного набора совпадений идентифицируются подмножества ключевых точек, которые согласуют объект и его местоположение, масштаб и ориентацию на новом изображении, чтобы отфильтровать хорошие совпадения. Определение последовательных кластеров выполняется быстро, используя эффективную реализацию хэш-таблицы обобщенного преобразования Хафа. Каждый кластер из трех или более объектов, которые согласуются с объектом и его позицией, затем подвергается дальнейшей детальной проверке модели. Впоследствии лишние объекты отбрасываются. Наконец, вычисляется вероятность того, что определенный набор признаков указывает на наличие объекта, с учетом точности соответствия и количества вероятных ложных совпадений.

Основным моментом в детектировании особых точек является построение пирамиды гауссианов и разностей гауссианов (Difference of Gaussian – DoG). Гауссианом (или изображением, размытым гауссовым фильтром) является изображение:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (2)$$

где L – значение гауссиана в точке с координатами (x, y) ; σ – радиус размытия; G – гауссово ядро; I – значение исходного изображения; $*$ – операция свертки [8].

Существуют и другие методы, основанные на описанных выше методах и расширяющие их возможности.

В *методе BRISK* (Binary Robust Invariant Scalable Keypoints) [9] обнаружение ключевой точки осуществляется на основании масштаба: точки интереса идентифицируются как по размеру изображения, так и по масштабу с использованием критерия значимости. Чтобы повысить эффективность вычислений, ключевые точки обнаруживаются в октавных слоях пирамиды изображения, а также в промежуточных слоях. Расположение и масштаб каждой ключевой точки могут быть получены в непрерывной области с помощью квадратичной функции.

Для дескриптора ключевой точки в данном методе используется образец выборки, состоящий из одной точки, лежащей на соответствующем образе масштабированных концентрических окружностях, который применяется в окрестности каждой ключевой точки для получения значений оттенков серого цвета: путем обработки локальных градиентов интенсивности и определения характерного направления функции. Ориентированный образец выборки BRISK используется для получения парных результатов сравнения яркостей, которые собираются в двоичный дескриптор BRISK.

Преимуществами *метода MSER* (Maximally Stable Extremal Regions) [10] являются:

- инвариантность к аффинному преобразованию интенсивностей изображения;
- ковариация к преобразованию (непрерывному) сохранения соседства $T : D \rightarrow D$ в области изображения;
- стабильность: выбираются только регионы, опора которых почти одинакова в диапазоне пороговых значений;
- многомасштабный поиск без какого-либо сглаживания: обнаруживается как

мелкая, так и большая структура;

- повторяемость: обнаружение MSER точек в масштабной пирамиде повышает повторяемость, а количество соответствий между масштабами изменяется;

- определенность: множество всех экстремальных областей можно перечислить в худшем случае $O(n)$, где n – количество пикселей на изображении.

Метод FREAK (Fast Retina Keypoint) [11] использует дескриптор ключевых точек, похожий на структуру зрительной системы человека, а точнее – сетчатки. Каскад двоичных строк вычисляется путем эффективного сравнения интенсивностей изображения по шаблону выборки сетчатки. Авторы утверждают, что FREAK точки обычно быстрее вычисляются с более низкой нагрузкой на память, а также они более надежны, чем SIFT, SURF или BRISK. Они предлагают рассматривать углы как ключевые точки. В данном дескрипторе используются углы Harris. Применяется сетка для отбора сетчатки круглой формы, отличающаяся более высокой плотностью точек вблизи центра. Предлагается подражать быстрым версиям движениям глаз, осуществляющимся скачками при зрительном поиске («саккадический поиск»), путем разбора дескриптора на несколько шагов. Применение метода начинается поиском с использованием первых 16 байтов дескриптора FREAK, представляющего необработанную информацию. Если расстояние меньше порога, продолжается сравнение со следующими байтами для анализа более точной информации. В результате выполняется каскад сравнений, ускоряющий еще один шаг согласования.

Метод KAZE Features [12] использует многомасштабный 2D-детектор и алгоритм описания в нелинейных масштабных пространствах. Предыдущие подходы обнаруживают и описывают признаки на разных уровнях масштаба путем построения или аппроксимации гауссовского масштабного пространственного изображения. Однако гауссовское размытие не учитывает естественные границы объектов и сглаживает в равной степени как детали, так и шум, снижая точность лока-

лизации и отличительность. Напротив, в данном методе обнаруживаются и описываются 2D-функции в пространстве нелинейного масштаба посредством нелинейной диффузионной фильтрации. Таким образом, становится возможным сделать размытие локально адаптивным к данным изображения, уменьшая шум, но сохраняя границы объектов, получая превосходную точность локализации и отличительность. Нелинейное масштабное пространство построено с использованием эффективных методов аддитивного разложения операторов (AOS) и диффузии переменной проводимости.

Метод FAST [13] применяется в тех случаях, когда в приложениях с частотой кадров в режиме реального времени используются функциональные точки и необходим высокоскоростной детектор признаков. Функциональные детекторы, такие как SIFT (DoG), Harris и SUSAN, являются хорошими методами, обеспечивающими высококачественные функции, однако они слишком требовательны к вычислительным ресурсам для использования в приложениях реального времени любой сложности. Авторы показывают, что машинное обучение может использоваться для получения детектора признаков, который может полностью обрабатывать видео в реальном времени с использованием менее 7 % доступного времени обработки. При сравнении с ранее описанными методами необходимо отметить, что ни детектор Харриса (120 %), ни SIFT (300 %) не могут работать с полной частотой кадров.

По мнению авторов [13], очевидным является вывод о том, что высокоскоростной детектор имеет ограниченное применение, если созданные признаки не подходят для последующей обработки. В частности, одна и та же сцена, просматриваемая с двух разных позиций, должна использовать признаки, соответствующие тем же реальным 3D-местоположениям. Поэтому вторым важным вкладом авторов являются угловые детекторы сравнения, основанные на этом критерии, применяемые к 3D-сценам.

Метод Harris [14] использует дискрет-



ные признаки изображения. По мнению авторов, чтобы обеспечить явное отслеживание признаков изображения, признаки изображения должны быть дискретными, а не формировать континуум, например, текстуру, или крайние пиксели (эглеры). Однако отсутствие возможности соединения точек-объектов является основным ограничением данного метода при получении дескрипторов более высокого уровня, таких как поверхности и объекты. В методе используются детекторы углов Моравеца, которые рассматривают локальное окно на изображении и определяют средние изменения интенсивности изображения, возникающие в результате сдвига окна на небольшое количество пикселей в разных направлениях.

Метод *MinEigen* [15] осуществляет выбор признаков и контроль функций во время отслеживания. Выбор специфически максимизирует качество отслеживания и поэтому оптимален по конструкции, в отличие от более специальных мер текстурирования. Мониторинг является недорогостоящим и надежным в вычислительной области и позволяет различать хорошие и плохие признаки на основе степени несходства, которая использует движение как базовую модель изменения изображения.

Представленные описания особенностей и функциональных возможностей методов обнаружения особых точек не позволяют сделать обоснованный выбор метода для формирования единого изображения

на основе ортографического вида сверху, полученного с четырех камер в разное время суток, поэтому основная задача экспериментального исследования — получение данных для выполнения сравнительного анализа методов обнаружения особых точек на изображениях при различных уровнях освещения.

Методика проведения экспериментов

Для проведения экспериментов использовался инструментарий среды моделирования **MATLAB**. Для повторяемости предлагаемых экспериментов и возможности верификации полученных результатов в программной реализации методов применялись встроенные функции определения особых точек (табл. 1) [16]. Для определения общих особых точек на двух изображениях использовалась функция `matchFeatures`.

Источниками исходных данных для экспериментальных исследований были несколько специализированных баз изображений (*Matlab Stitching Example* [17], *Adobe Panoramas Dataset* [18], *Panoramic Image Database* [19], *PIA Panorama Dataset* [20]), содержащих в совокупности 445 наборов для панорам, в каждый из которых входило от 5 до 32 изображений (рис. 1, 2), которые в дальнейшем сшивались в единое изображение по особым точкам, полученным разными методами.

Для проведения экспериментов изображения подвергались искусственному затемнению, для чего использовалась встроенная функция `imadjust` среды **MATLAB** [21].

Таблица 1

Функции поиска особых точек

Функция	Тип особых точек
<code>detectFASTFeatures</code>	Углы. Одномасштабное обнаружение
<code>detectMinEigenFeatures</code>	
<code>detectHarrisFeatures</code>	
<code>detectBRISKFeatures</code>	Углы. Многомасштабное обнаружение
<code>detectSURFFeatures</code>	«Капли» (blobs). Многомасштабное обнаружение
<code>detectMSERFeatures</code>	Области однородной интенсивности. Многомасштабное обнаружение



Рис. 1. Пример исходного набора изображений (Matlab Stitching Example)

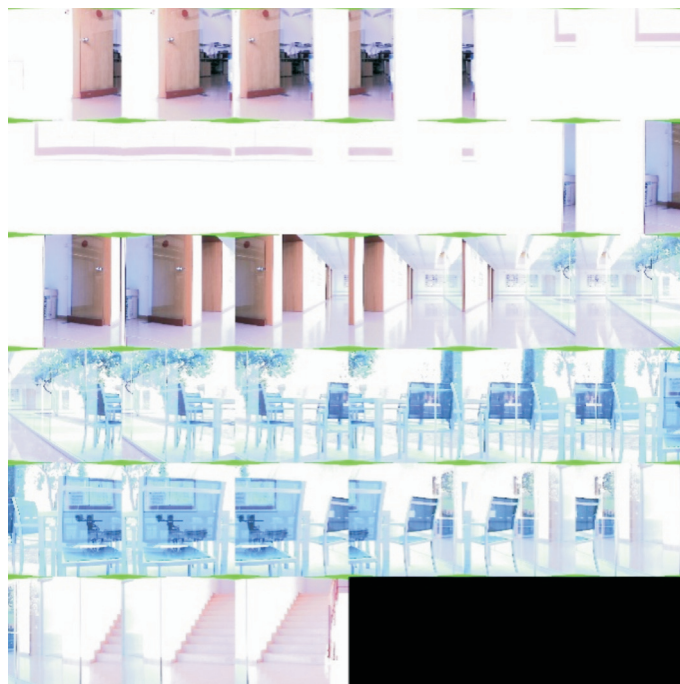


Рис. 2. Пример исходного набора изображений
(cvc01passadis-cyl-pano10 из IIIA Panorama dataset)



Рис. 3. Набор изображений после последовательного затемнения (слева направо — исходное изображение, $\gamma = 2$, $\gamma = 4$)

Весь диапазон значений яркости изображения (0; 255) преобразовывался в диапазон (0; 150), коэффициент гамма-коррекции γ при этом последовательно менялся от 1 до 6 с шагом 0.1, что позволило получить по 50 модифицированных изображений для каждого набора (рис. 3).

В соответствии с табл. 1 каждый из наборов подвергался обработке в MATLAB для получения единого изображения. На рис. 4 показаны изображения, полученные при сшивке по особым точкам, которые определены методом SURF.

Проведение визуального сравнительного анализа эффективности обработки методами SURF и MSER из набора 4 (рис. 5) по итоговому изображению панорамы достаточно затруднительно. Для выполнения оценки эффективности методов предложено выполнение сравнения результатов осуществлять на более раннем этапе работы алгоритмов: на этапе определения общих особых точек на двух изображениях. Метрикой для сравнения методов выбрана оценка среднего количества общих особых точек на изображе-



Рис. 4. Результаты сшивки наборов изображений с разным значением γ методом SURF

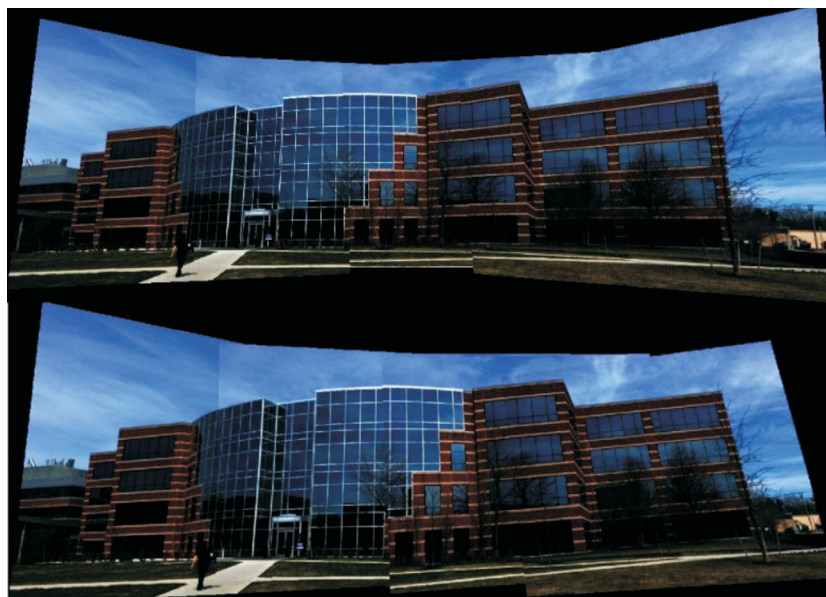


Рис. 5. Сшивки изображений с поиском особых точек разными методами:
сверху — методом SURF, снизу — методом MSER

ниях для каждого набора. Результаты расчета среднего количества общих особых точек на изображениях для каждого набора для разных методов представлены в следующем разделе статьи.

Результаты

В ходе проведения экспериментальных исследований качество сшивки изображений определялось по количеству общих точек между парой анализируемых изображений.

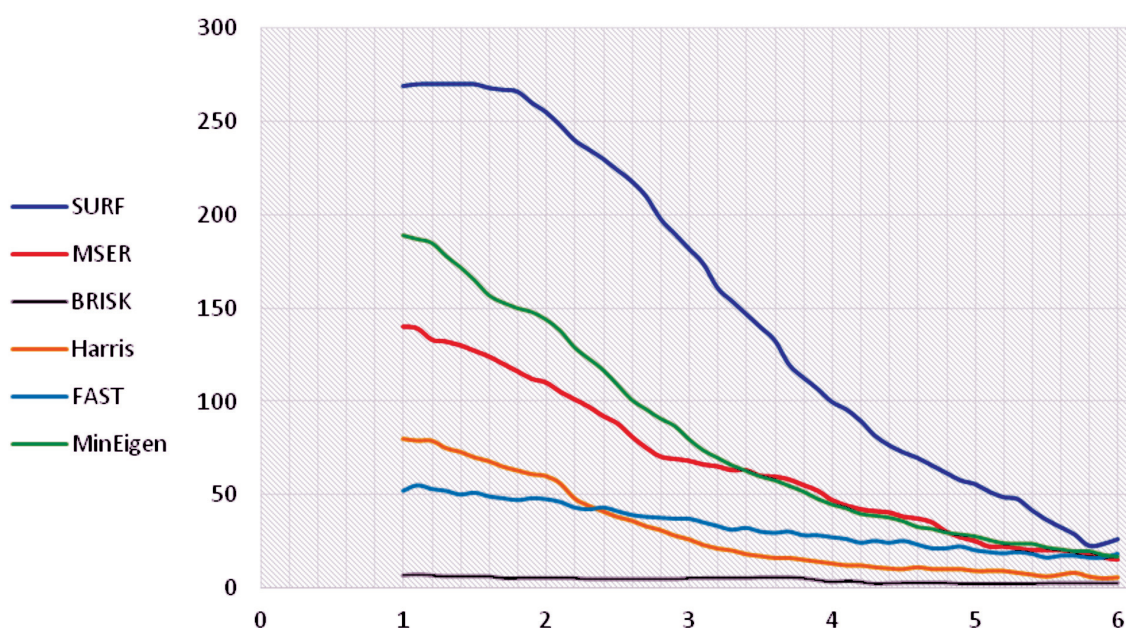


Рис. 6. График зависимости количества общих особых точек от уровня освещенности изображения



Таблица 2

Значения инвариантности методов поиска особых точек
к степени освещения

Метод	Инвариантность, %
SURF	9,85
MSER	10,71
BRISK	34,62
Harris	6,88
FAST	34,62
MinEigen	8,99

На графике (рис. 6) показана зависимость количества общих особых точек от уровня освещенности.

График показывает, что метод SURF имеет наилучшие результаты как при сильном, так и при слабом освещении, в том числе при съемке ночью.

Другой важный параметр, по которому нами сравнивались методы поиска особых точек, — это инвариантность к степени освещения, рассчитываемая по формуле:

$$inv = \frac{\min}{\max} * 100 \%, \quad (3)$$

где $\max$ — максимальное количество общих точек между парой изображений; $\min$ — минимальное количество общих точек между парой изображений.

В табл. 2 приведены значения инвариантности для всех методов.

В соответствии с полученными результатами можно сделать вывод, что метод SURF не инвариантен к изменению освещения, а, следовательно, качество сшивки с его применением будет значительно отличаться в светлое и темное время суток.

Методы BRISK и FAST показали одинаково высокую инвариантность к освещению, т. е. оно наименее значительно повлияло на качество поиска особых точек и, соответственно, на сшивку. Значительным минусом метода BRISK при этом является очень маленькое количество най-

денных общих особых точек для пары изображений.

Другой интересный вывод, который можно сделать из приведенного выше графика: методы MinEigen и MSER значительно превосходят FAST при высоком освещении, при низком уровне FAST показывает результаты немного лучше конкурентов, что говорит о том, что FAST не только инвариантен, но и в целом лучше работает при низком освещении.

Заключение

Таким образом, нами проведены эксперименты по нахождению особых точек на изображениях методами SURF, MSER, BRISK, Harris, FAST и MinEigen. Реализован поиск общих особых точек для пары изображений, а также анализ их количества. Проведена сшивка изображений разными методами при различных уровнях освещений.

Результаты показали, что из исследованных нами методов SURF является наилучшим по качеству полученной сшивки, а метод FAST наиболее инвариантен к изменению освещения.

В дальнейшем предполагается провести сравнение этих методов по их быстродействию, что необходимо для сшивки в режиме real time. Тогда на основании данного и следующего исследований можно будет сделать выводы о том, какие методы, в какой комбинации и при каких условиях

следует применять. Планируется разработка нового метода, который должен иметь высокое качество сшивки, инвариантность

к освещению и быстродействие, позволяющее производить сшивку изображений в реальном времени.

СПИСОК ЛИТЕРАТУРЫ

1. Почему автомобили попадают в аварии? // URL: <http://www.transportal.by/obsujdaem/pochemu-gruzovye-avtomobilipopadayut-v-avarii.html> (Дата обращения: 23.09.2017).
2. Что такое мертвая зона и как обезопаситься? // URL: <http://www.gazu.ru/safety/bdd/10225/> (Дата обращения: 23.09.2017).
3. **Makarov A.S., Bolsunovskaya M.V.** The 360° around view system for large vehicles, the methods of calibration and removal of barrel distortion for omnidirectional cameras // Supplementary Proc. of the 5th Internat. Conf. on Analysis of Images, Social Networks and Texts. Yekaterinburg, 2016. Vol. 1710. Pp. 183–192.
4. **Makarov A.S., Bolsunovskaya M.V.** The method of obtaining orthographic top view for around view system at vehicles // Proc. of the 2017 IEEE Conf. of Russian Young Researchers in Electrical and Electronic Engineering. St. Petersburg. IEEE, 2017. Pp. 697–699.
5. **Bay H., Ess A., Tuytelaars T., Gool L.V.** Speeded-up robust features (SURF) // Computer Vision and Image Understanding. Academic Press, 2008. Vol. 110. No. 3. Pp. 346–359.
6. Применение метода SURF в системах контроля и управления доступом на основе биометрических технологий // URL: <https://habrahabr.ru/post/152679/> (Дата обращения: 23.09.2017).
7. **Lowe D.G.** Object recognition from local scale-invariant features // The Proc. of the 7th IEEE Internat. Conf. on Computer Vision. 1999. Vol. 2. Pp. 1150–1157.
8. Построение SIFT дескрипторов и задача сопоставления изображений // URL: <https://habrahabr.ru/post/106302/> (Дата обращения: 23.09.2017).
9. **Leutenegger S., Chli M., Siegwart R.Y.** BRISK: Binary robust invariant scalable keypoints // The Proc. of the IEEE Internat. Conf. on Computer Vision. 2011. Pp. 2548–2555.
10. **Matas J., et al.** Robust wide-baseline stereo from maximally stable extremal regions // Image and Vision Computing. 2004. Vol. 22. No. 10. Pp. 761–767.
11. **Alahi A., Ortiz R., Vandergheynst P.** Freak: Fast retina keypoint // Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. 2012. Pp. 510–517.
12. **Alcantarilla P.F., Bartoli A., Davison A.J.** KAZE features // European Conf. on Computer Vision. Springer, Berlin, Heidelberg, 2012. Pp. 214–227.
13. **Rosten E., Drummond T.** Machine learning for high-speed corner detection // European Conf. on Computer Vision. 2006. Pp. 430–443.
14. **Harris C., Stephens M.** A combined corner and edge detector // Alvey Vision Conference. 1988. Vol. 15. No. 50. Pp. 105–244.
15. **Shi J., et al.** Good features to track // Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. 1994. Pp. 593–600.
16. Point Feature Types // URL: <https://www.mathworks.com/help/vision/ug/point-feature-types.html> (Дата обращения: 23.09.2017).
17. Feature Based Panoramic Image Stitching // URL: <https://www.mathworks.com/help/vision/examples/feature-based-panoramic-image-stitching.html>
18. Adobe Panoramas Dataset // URL: <https://sourceforge.net/adobe/adobedatasets/panoramas/home/Home/> (Дата обращения: 31.03.2018).
19. **Muller R., Vardy A., Kreft S., Ruwisch S.** Visual homing in environments with anisotropic landmark distribution // Autonomous Robots. 2007. Vol. 23(3). Pp. 231–245 // URL: <http://www.ti.uni-bielefeld.de/html/research/avardy/index.html> (Дата обращения: 31.03.2018).
20. **Ramisa A., Tapus A., Lopez de Mantaras R., Toledo R.** Mobile robot localization using panoramic vision and combination of local feature region detectors // Proc. of the IEEE Internat. Conf. on Robotics and Automation. Pasadena, California, 2008. Pp. 538–543 // URL: <http://www.iiia.csic.es/~aramisa/datasets/iiapanos.html> (Дата обращения: 31.03.2018).
21. Список функций Image Processing Toolbox: Улучшение изображений. Imadjust // URL: <http://matlab.exponenta.ru/imageprocess/book3/10/imadjust.php> (Дата обращения: 31.03.2018).

Статья поступила в редакцию 10.11.2017.

REFERENCES

1. *Pochemu avtomobili popadayut v avarii?* [Why do cars get into accidents?] Available: <http://www.transportal.by/obsujdaem/pochemu-gruzovye-avtomobilipopadayut-v-avarii.html> (Accessed: 23.09.2017). (rus)
2. *Chto takoye mertvaya zona i kak obezopasitsya?* [What is a dead zone and how to protect yourself?] Available: <http://www.gazu.ru/safety/bdd/10225/> (Accessed: 23.09.2017). (rus)
3. **Makarov A.S., Bolsunovskaya M.V.** The 360° around view system for large vehicles, the methods of calibration and removal of barrel distortion for omnidirectional cameras. *Supplementary Proceedings of the 5th International Conference on Analysis of Images, Social Networks and Texts*, Yekaterinburg, Russia, 2016, Vol. 1710, Pp. 183–192.
4. **Makarov A.S., Bolsunovskaya M.V.** The method of obtaining orthographic top view for around view system at vehicles. *Proceedings of the 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering*, St. Petersburg, Russia, 2017. IEEE, 2017, Pp. 697–699.
5. **Bay H., Ess A., Tuytelaars T., Gool L.V.** Speeded-up robust features (SURF). *Computer Vision and Image Understanding*. Academic Press, 2008, Vol. 110, No. 3, Pp. 346–359.
6. *Primeneniye metoda SURF v sistemakh kontrolya i upravleniya dostupom na osnove biometricheskikh tekhnologiy* [Application of the SURF method in control and management systems based on biometric technologies]. Available: <https://habrahabr.ru/post/152679/> (Accessed: 23.09.2017). (rus)
7. **Lowe D.G.** Object recognition from local scale-invariant features. *Proceedings of the 7th IEEE International Conference on Computer Vision*. IEEE, 1999, Vol. 2, Pp. 1150–1157.
8. *Postroyeniye SIFT deskriptorov i zadacha sopostavleniya izobrazheniy* [The construction of SIFT descriptors and the task of mapping images]. Available: <https://habrahabr.ru/post/106302/> (Accessed: 23.09.2017).
9. **Leutenegger S., Chli M., Siegwart R.Y.** BRISK: Binary robust invariant scalable keypoints. *Proc. of the IEEE International Conference on Computer Vision*. IEEE, 2011, Pp. 2548–2555.
10. **Matas J., et al.** Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 2004, Vol. 22, No. 10, Pp. 761–767.
11. **Alahi A., Ortiz R., Vandergheynst P.** Freak: Fast retina keypoint. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, Pp. 510–517.
12. **Alcantarilla P.F., Bartoli A., Davison A.J.** KAZE features. *European Conference on Computer Vision*. Springer, Berlin, Heidelberg, 2012, Pp. 214–227.
13. **Rosten E., Drummond T.** Machine learning for high-speed corner detection. *European Conference on Computer Vision*, 2006, Pp. 430–443.
14. **Harris C., Stephens M.** A combined corner and edge detector. *Alvey Vision Conference*, 1988, Vol. 15, No. 50, Pp. 105–244.
15. **Shi J., et al.** Good features to track. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 1994, Pp. 593–600.
16. *Point Feature Types*. Available: <https://www.mathworks.com/help/vision/ug/point-feature-types.html> (Accessed: 23.09.2017).
17. *Feature Based Panoramic Image Stitching*. Available: <https://www.mathworks.com/help/vision/examples/feature-based-panoramic-image-stitching.html>
18. *Adobe Panoramas Dataset*. Available: <https://sourceforge.net/adobe/adobedatasets/panoramas/home/Home/> (Accessed: 31.03.2018).
19. **Müller R., Vardy A., Kreft S., Ruwisch S.** Visual homing in environments with anisotropic landmark distribution. *Autonomous Robots*, 2007, Vol. 23(3), Pp. 231–245. Available: <http://www.ti.uni-bielefeld.de/html/research/avardy/index.html> (Accessed: 31.03.2018).
20. **Ramisa A., Tapus A., Lopez de Mantaras R., Toledo R.** Mobile robot localization using panoramic vision and combination of local feature region detectors. *Proceedings of the IEEE International Conference on Robotics and Automation*, Pasadena, California, 2008, Pp. 538–543. Available: <http://www.iiia.csic.es/~aramisa/datasets/iiapanos.html> (Accessed: 31.03.2018).
21. *Spisok funktsiy Image Processing Toolbox: Uluchsheniye izobrazheniy. Imadjust* [List of Functions Image Processing Toolbox: Image enhancement. Imadjust]. Available: <http://matlab.exponenta.ru/imageprocess/book3/10/imadjust.php> (Accessed: 31.03.2018). (rus)

Received 10.11.2017.

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

МАКАРОВ Алексей Сергеевич
MAKAROV Aleksei S.
E-mail: lyohamakarov@yandex.ru

БОЛСУНОВСКАЯ Марина Владимировна
BOLSUNOVSKAYA Marina V.
E-mail: bolsun_hht@mail.ru

РАЗРАБОТКА И ИССЛЕДОВАНИЕ СИСТЕМЫ ПРЯМЫХ СОЕДИНЕНИЙ D2D С СОТОВОЙ ПОДДЕРЖКОЙ LTE

А.В. Пяттаев¹, С.Д. Андреев¹, Е.А. Кучерявый²

¹ Российский университет дружбы народов,
Москва, Российская Федерация;

² Национальный исследовательский университет «Высшая школа экономики»,
Москва, Российская Федерация

Ввиду популярности мобильных пользовательских приложений операторы сотовой связи все чаще сталкиваются с задачей разгрузки своего ограниченного лицензированного спектрального ресурса. Применение для этих целей технологий радиодоступа, функционирующих в нелицензированном спектре, является привлекательным для сокращения соответствующих затрат. В статье описана система прямых соединений D2D в нелицензированной полосе частот на основе технологии WiFi Direct для разгрузки мобильного трафика, обслуживаемого сотовой системой LTE. Разработан протокол обеспечения сетевой поддержки соединений D2D со стороны архитектуры сети LTE, а также произведено имитационное моделирование на системном уровне для выявления соответствующего повышения эффективности управления трафиком. Описаны испытания связи D2D на модельной сети LTE.

Ключевые слова: мобильный трафик, сотовая связь LTE, прямые соединения с поддержкой сети, технология WiFi Direct, исследование и внедрение.

Ссылка при цитировании: Пяттаев А.В., Андреев С.Д., Кучерявый Е.А. Разработка и исследование системы прямых соединений D2D с сотовой поддержкой LTE // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2018. Т. 11. № 2. С. 19–34. DOI: 10.18721/JCSTCS.11202

DESIGN AND EVALUATION OF A SYSTEM OF DIRECT D2D CONNECTIONS WITH LTE CELLULAR ASSISTANCE

A.V. Pyattaev¹, S.D. Andreev¹, Ye.A. Koucheryavy²

¹ Peoples' Friendship University of Russia (RUDN University), Moscow, Russian Federation;

² National Research University Higher School of Economics, Moscow, Russian Federation

Due to the growing popularity of mobile user applications, cellular network operators are increasingly interested in offloading their scarce licensed spectrum resources. To this aim, using unlicensed-band radio technologies is attractive owing to its reduced costs. This work designs a system of direct D2D (device-to-device)

connections to offload mobile traffic that is served by LTE (long-term evolution) cellular networks. In particular, a network assistance protocol over LTE architecture is proposed for D2D connections, and its system-level performance is evaluated to quantify the resulting traffic-steering efficiency. In addition, practical operation of D2D communication in a test LTE network is described, which allows making substantiated conclusions about the actual feasibility of the proposed solution in real-life conditions.

Keywords: mobile traffic, LTE cellular communication, direct connections with network assistance, WiFi Direct radio technology, design and evaluation.

Citation: Pyattaev A.V., Andreev S.D., Koucheryavy Ye.A. Design and evaluation of a system of direct D2D connections with LTE cellular assistance. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems, 2018, Vol. 11, No. 2, Pp. 19–34. DOI: 10.18721/JCSTCS.11202

Введение и обзор литературы

Лицензированный беспроводной спектр продолжает оставаться крайне ограниченным и дорогим. Поскольку имеющегося спектра уже недостаточно для удовлетворения существующего спроса со стороны операторов мобильной связи, предпринимаются попытки лицензировать дополнительный спектральный ресурс [1]. Поэтому использование прямых соединений между пользовательскими устройствами (device-to-device – D2D) в нелицензированной полосе частот для целей выгрузки мобильного трафика исключительно важная задача. При этом нелицензированные спектральные ресурсы не могут быть зарезервированы только для целей D2D и, следовательно, соединения D2D должны использовать их совместно с существующими услугами беспроводного доступа.

Расширенное управление интерференцией с контролем доступа и мощности [2] требуется для поддержки нескольких соединений D2D на одном и том же канале. Исследовательские работы (например, [3–5]) предлагают способ контроля интерференции, использующий информацию о состоянии канала в реальном времени, местоположении пользователя, а также о состоянии среды передачи. При этом схема управления интерференцией должна решить, как поступить с каждым соединением D2D:

разделять общие ресурсы лицензированной полосы со стандартными сотовыми передачами (между пользователями и базовыми станциями) и испытывать интерференцию;

использовать выделенные ресурсы и стремиться обеспечить их эффективное использование;

оставаться в рамках сотовой инфраструктуры.

Данный процесс, известный также как *выбор режима*, привлек внимание исследователей, сфокусировавшихся на широком круге оптимизационных задач для таких показателей, как отношение сигнала и помехи к шуму (SINR) [6], пропускная способность [7], энергетическая эффективность [8], задержка передачи данных [9], справедливое распределение ресурсов и вероятность отказа [10]. Большинство этих исследований рассматривало D2D-технологии в лицензированном спектре в системах долгосрочной эволюции 3GPP LTE (long-term evolution) [11, 12].

С другой стороны, оператор не может обеспечить эксклюзивного использования какого-либо нелицензированного спектра. В результате интерференция становится непредсказуемой, что требует надежного контроля доступа к среде, способного справиться с данной проблемой. Например, были разработаны технологии Bluetooth и WiFi, которые становятся все более популярными в беспроводных персональных и локальных сетях (WPAN/WLAN). Основанное на стандартах IEEE 802.11 решение WiFi в настоящее время является преобладающим для подключения пользовательских устройств [13].

Однако поскольку WiFi не имеет гаран-

тированного контроля доступа, эту технологию часто критикуют за то, что она не всегда удовлетворяет требованиям по качеству обслуживания QoS [14]. Тем не менее, WiFi обеспечивает более высокие скорости передачи данных и энергоэффективность, чем любая из сотовых технологий. В принципе, при наличии соответствующих методов управления, для WiFi-соединения можно добиться стабильных результатов работы [15]. С увеличением количества соединений D2D и персональных сетей, в том числе построенных на более современной технологии WiFi Direct, такое управление становится необходимым.

В результате сотовые сети могут обеспечить интеллектуальное управление радиоресурсами, а также столь необходимую анонимность при обслуживании. Если пользовательские устройства постоянно связаны с сотовой сетью, она может предложить резервное соединение, когда для нелицензированной передачи QoS оказывается неудовлетворительным. Кроме того, при поддержке со стороны сотовой сети аутентификация устройств и безопасность линии D2D могут быть значительно улучшены, что, например, делает любой вид атаки «человек посередине» практически невозможным. Другими словами, наличие поддержки со стороны сотовой сети для D2D-соединений на нелицензированных частотах так, как это предлагается, например, в [16], представляется крайне перспективным подходом.

Архитектурная реализация системы прямых соединений

Как было сказано выше, для сотовых операторов желательной является выгрузка мобильного трафика в сеть D2D нелицензированного спектра как конкурентное решение с LTE- и WiFi-линиями для мультирадиоустройств. При этом ни WiFi, ни WiFi Direct (WFD) не располагают быстрыми и эффективными методами обнаружения устройств/услуг, а также не имеется простого способа внедрения D2D-взаимодействий вследствие очень коротких расстояний между устройствами. Наконец, анонимность прямых соединений остается

открытой задачей для решений D2D в нелицензированном спектре. С целью преодоления этих ограничений далее предлагается новая архитектура для внедрения поддерживаемых сетью D2D-решений.

Схема работы при обслуживании соединений D2D. Для начала определим, какие именно услуги должны быть обеспечены для конечного пользователя. Естественно, очень сложно гарантировать какие-либо адекватные условия передачи для D2D-линий, а качество обслуживания для таких соединений может существенно отличаться в различные моменты времени и в связи с перемещением пользователей. Поэтому услуги, толерантные к задержкам, такие как распределенное кэширование и кооперативная передача, могут рассматриваться как первые кандидаты на выгрузку трафика. Однако наиболее интересными для мобильных пользователей представляются услуги по запросу, требующие стабильных линий высокого качества и, таким образом, никогда не предполагавшиеся для случая использования D2D-технологий.

Тем не менее, если устройства находятся близко друг к другу и можно предсказать достаточно стабильный уровень качества связывающей их линии, многие запрашиваемые услуги становятся реализуемыми. На коротком промежутке времени работы приложения, такие как потоковое видео, игры со многими участниками и т. п., могут быть предоставлены посредством D2D-линий. Чтобы сделать эти услуги коммерчески привлекательными, необходимы определенные гарантии QoS, а также способы определения того, в какой мере выполнены предварительные условия для использования D2D.

Поэтому критически важные требования к разработке архитектуры сети D2D состоят в предоставлении пользователям знания о том, когда можно устанавливать D2D-соединения, с кем, какой контент доступен, и что делать, если почему-то нет возможности поддерживать QoS. Для этого, в частности, должна быть предусмотрена поддержка следующих функций:

- идентификации цели взаимодействия (хранение и поиск файловых и потоковых

данных, идентификация участников игры и т. д.);

- поиска ближайших партнеров, проявляющих интерес к контенту;
- аутентификации и авторизации для целей передачи контента;
- поддержки *надежного* дублирования при отказе D2D-линии.

Технически нахождение пары, имеющей подходящий контент, не является сложной задачей вследствие широкого распространения облачных и социальных сетей. Важнейшая проблема состоит в том, что соединение, которое еще не установлено, не может быть представлено традиционным образом: нет ассоциированных с ним интерфейсов, нет IP-адресов и т. д. Естественный выход из такой ситуации — необходимость создания нового объекта, специально разработанного для управления D2D-линией и принятия решений в реальном времени, основанных на позиционировании и доступности радиоресурсов. Поскольку такая информация в системе управления связью собирается только оператором, этот объект должен быть частью операторской сети.

С другой стороны, программные интерфейсы для приложений (API) доступны только провайдеру услуг. Тогда провайдер и оператор должны организовать взаимодействие для предоставления комплексных услуг пользователю. Это наиболее естественный интерфейс между двумя различными технологиями: специфичными возможностями контроля линии (управляемой по запросу сервером D2D, являющимся частью транспортной сети) и специфическими возможностями отслеживания контента (управляемого по запросу сервером приложений). Предложенное в данной статье решение (рис. 1) функционирует следующим образом:

1. Каждое пользовательское устройство применяет возможности уровня приложений для аутентификации с сервером приложений (например, Facebook). Это позволяет обеспечить взаимодействие с контентом, который авторизуется третьей стороной, а также устраняет требования по обработке метаданных и предоставляет возможность авторизации со стороны социальных сетей.

2. Пользовательское оборудование (user equipment — UE) позволяет серверу D2D своего оператора представлять его на сервере приложений при организации D2D-соединений. При этом сервер D2D никогда не получает доступ к контенту или его метаданным: осуществляется лишь аутентификация пользователя, чтобы убедиться, что устройство действительно принадлежит владельцу заявленных учетных данных уровня приложения. Таким образом, сервер D2D позволяет присваивать имена прикладного уровня физическим устройствам, в том числе и в случае, когда имеется несколько устройств, принадлежащих одному и тому же лицу.

3. Пользовательское устройство может публиковать или искать линии для передачи контента на сервере приложений, и эти линии будут ссылаться на конкретный контент некоторого пользователя (но не устройства). Линии могут быть постоянными, ограниченными по времени или одно-разовыми в зависимости от целей приложения.

4. Пользователь запрашивает сервер D2D перед созданием прямого соединения, чтобы обеспечить линию на уровне приложений, а сервер приложений преобразует этот запрос в фактическое соединение на уровне линии и обеспечивает IP-адрес, к которому могут быть привязаны сокет.

5. Наконец, может производиться обмен данными. Заметим, что сервер приложений не участвует в этом процессе и не отслеживает обмен контентом непосредственно, лишь удостовераясь в том, что линии следуют модели безопасности, не принимая участия в их локальном управлении. Сервер D2D при этом может осуществлять мониторинг и регулировать свойства D2D-линии при необходимости и в зависимости от его функций.

Далее кратко рассмотрим те особенности предложенной схемы, которые делают ее приемлемым решением с точки зрения безопасности для поддерживаемой сотовой сетью связи D2D.

Каждое соединение начинается с проверки контента линии. Потенциальным злоумышленникам необходимо будет по-

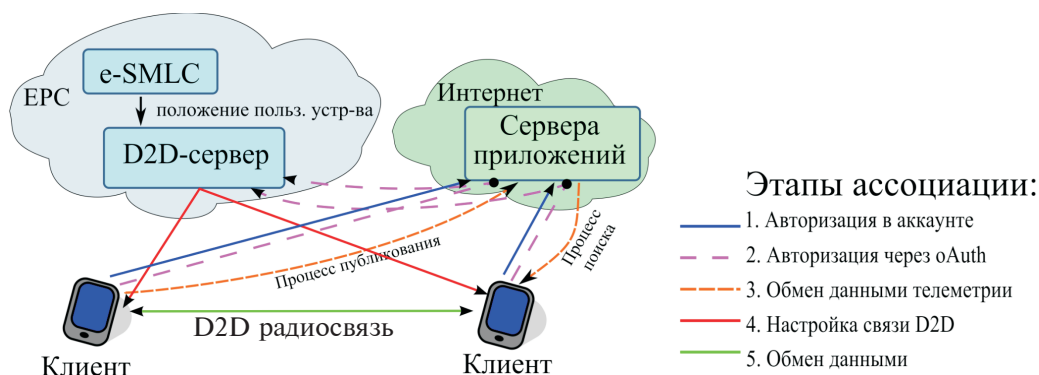


Рис. 1. Поддержка сети при создании D2D-линии

лучить доступ к серверу приложений для получения свойств такой линии. Поэтому любая атака на систему по существу блокируется одноразовым паролем, распространяемым сервером приложений через существующие защищенные каналы.

Ни одно из пользовательских устройств не должно широкоэмитивно транслировать никакую информацию об обнаружении или прослушивать запросы на обнаружение. Поскольку устройства поддерживают свои

радиоинтерфейсы D2D выключенными до тех пор, пока не будет получена соответствующая команда активации, они не могут быть атакованы или отслежены до того, как перейдут в активное состояние.

Интеграция с сотовой архитектурой 3GPP. Подчеркнем, что предложенный выше протокол является рабочим решением и не претендует на оптимальность. Он в значительной степени опирается на существующие решения, такие как OAuth и

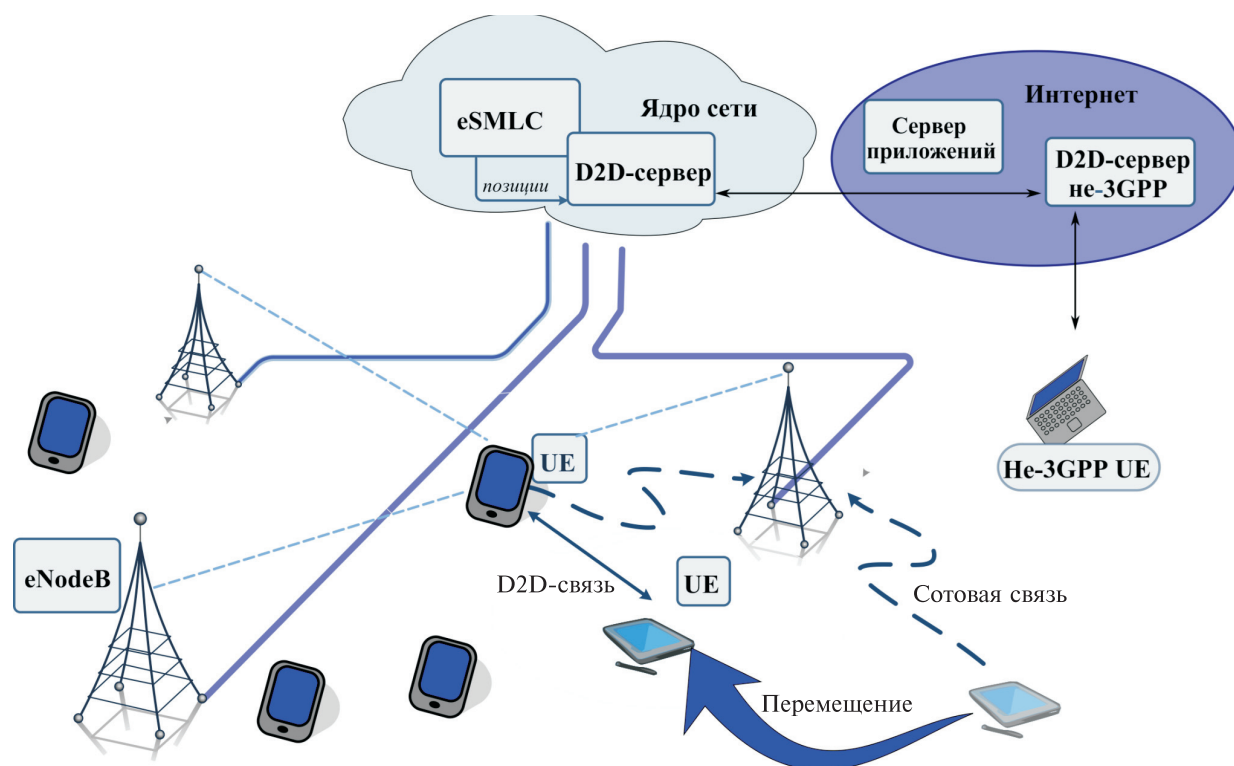


Рис. 2. Предложенное размещение сервера D2D в архитектуре 3GPP

SSL, для авторизации на сервере D2D при управлении соединениями и передаче команд, соответственно. Кроме того, протокол предполагает, что серверу D2D всегда можно доверять только потому, что он находится в сети оператора. Поэтому на практике необходимо оценить потенциальные риски безопасности в конкретных случаях, когда применяется связь D2D с поддержкой со стороны сотовой сети.

Далее сосредоточимся на вопросах интеграции системы с существующими сетевыми архитектурами. Благодаря гибкости ядра сети 3GPP относительно легко выполнить интеграцию сервера D2D в существующую архитектуру 3GPP LTE. На рис. 2 показано, как подобная интеграция может быть осуществлена.

Предложенная в данной статье схема может быть сопоставлена с объектами архитектуры 3GPP практически полностью, с единственным дополнительным объектом – сервером D2D, находящимся в EPC-фрагменте сети. Такое размещение позволяет серверу D2D связываться с центром определения местоположения (SMLC) для позиционирования пользовательского устройства, а также эффективно взаимодействовать с серверами приложений внешних провайдеров и операторов.

Фактически протокол для выгрузки трафика на соединения D2D на сегодняшний день уже является частью стандартов 3GPP. При этом некоторые детали стандартизованного решения 3GPP отличаются от первоначально предложенных нами. В частности, поток сообщений, описанный в документах 3GPP, не учитывает пробле-

мы взаимодействия с серверами приложений, поскольку это выходит за рамки работы 3GPP.

Исследование выгрузки трафика на прямые соединения

В данном разделе вводятся сетевые объекты и соответствующие механизмы, которые требуются для поддерживаемой со стороны сотовой сети технологии WiFi Direct. Далее предлагается описание разработанной системы имитационного моделирования, которая может использоваться для широкого класса перспективных приложений D2D и моделей трафика.

Подход к анализу эффективности работы.

В настоящей статье беспроводная сеть представляется как система, состоящая из ряда объектов, имеющих возможность поддерживать различные радиотехнологии (рис. 3).

Рассмотрим сеть 3GPP LTE, представленную базовыми станциями E-UTRAN NodeB (eNB). Каждая базовая станция связана с ядром сети, обеспечивающим сотовый доступ для всех беспроводных устройств, ассоциированных с сетью.

Любая базовая станция eNB доступна ряду мультирадиопользовательских устройств, каждое из которых имеет возможность взаимодействовать как с LTE, так и с WiFi. Каждому клиенту доступны возможности управления, реализуемые координатором уровня доступа (MAC), представляющим собой коммутатор потоков трафика. В зависимости от рекомендаций координатора, клиент может воспользоваться для передачи данных интерфейсом LTE или WiFi.

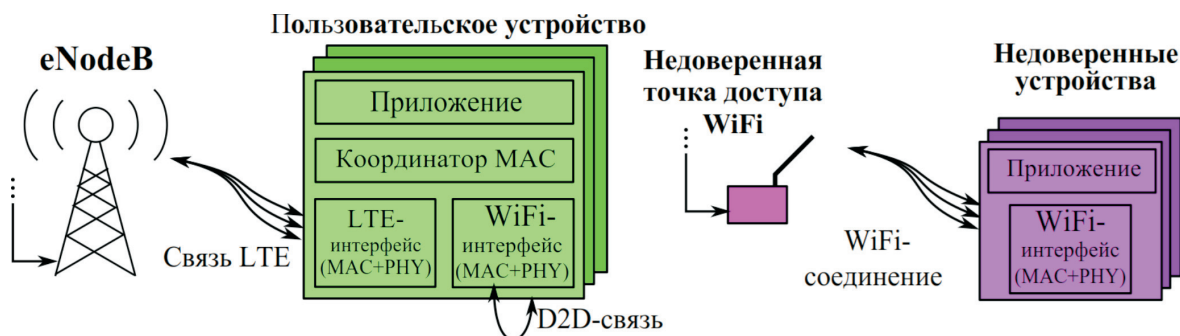


Рис. 3. Диаграмма объектов рассматриваемой сети

Необходимо также принимать во внимание интерференцию от устройств WLAN в нелицензированном спектре, которые могут быть расположены в непосредственной близости от точек доступа WiFi. В соответствии с принципами построения сетей WiFi, эти устройства имеют равные права с мультирадиоклиентами сети с точки зрения доступа к канальным ресурсам. В то же время они не могут быть ассоциированы с сотовой сетью, а значит их функционирование на нелицензированных частотах не может контролироваться или отслеживаться сетью LTE. Они именуются здесь «блуждающими» устройствами.

В соответствии с методологией исследований, заданной рекомендациям 3GPP в [17], предположим, что N клиентов LTE (которые в рассматриваемом случае имеют также возможность использования связи D2D) расположены равномерно в зоне покрытия сети. Плотность устройств выбирается достаточно высокой, так что каждый клиент в зоне действия связи D2D имеет возможность прямого соединения по крайней мере еще с одним клиентом. Однако только половина устройств может служить источниками данных и передавать трафик, поскольку прямые взаимодействия являются однонаправленными. При этом для каждой сессии существует только один источник и один получатель данных.

Такой трафик моделируется как насыщенный буфер с пакетами длиной 1500 байт каждый. Далее вместо моделирования некоторого распределения контента и запросов клиентов предполагается, что определенный процент устройств-источников x в радиусе действия D2D соединяется с соответствующими устройствами-получателями данных. Блуждающие устройства, предоставляющие другим пользователям «тяжелый» контент, также представляются моделью заполненного буфера с пакетами длиной 1500 байт, направляемыми к своей точке доступа. Для моделирования нисходящего трафика откорректируем число блуждающих устройств, чтобы получить некоторый уровень конкуренции на нелицензированных частотах.

Имитационное моделирование системы D2D. Для проведения исследований рассматриваемой системы разработан имитатор системного уровня (system-level simulator — SLS), базирующийся на методологии LTE, приведенной в TR 36.814 [17] и других документах 3GPP, а также в спецификациях IEEE для оценки производительности беспроводных сетей [18] и в текущих стандартах IEEE 802.11. Имитатор представляет собой гибкий инструмент, который может поддерживать различные способы построения сетей, модели трафика, характеристики каналов и беспроводные протоколы. Он позволяет моделировать существующую инфраструктуру системы LTE и требуемое размещение клиентов.

Для каждого устройства предусмотрен собственный выделенный генератор трафика, позволяющий варьировать его показатели. Каналы моделируются с учетом соответствующих характеристик источника, получателя и среды распространения. Каждый узел способен поддерживать несколько радиointерфейсов, которые могут использоваться для восходящих и нисходящих соединений и моделируются при этом в пределах единой временной шкалы. Этот имитатор был изначально предназначен для широкого круга задач и может рассматриваться как расширяемая система, подходящая для поддержки различных сценариев D2D и сотового взаимодействия.

Для оценки характеристик WiFi Direct с поддержкой от сотовой сети рассмотрим следующий сценарий, основанный на рекомендациях 3GPP при развертывании сетей в городских условиях. Сеть LTE включает в себя 19 шестиугольных сот с тремя секторами в каждой. Базовые станции eNB поддерживают технологические решения 3GPP LTE, расстояние между близлежащими базовыми станциями составляет 200 м, а радиус соты равен примерно 110 м. Все соты работают в одном и том же лицензированном спектре с полосой частот 60 МГц, которая разделена на три пары полос по 10 МГц для функционирования в режиме FDD.

Каждая сота, в свою очередь, разделена

на три сектора, и каждому сектору выделена пара полос по 10 МГц, что приводит к шаблону повторного использования частот 1х3х3. Клиенты 3GPP LTE связываются с базовой станцией eNB на основе наилучшего значения DL SINR с порогом переключения в 1 дБ. Для получения дополнительной информации о стандартной конфигурации LTE можно обратиться к табл. 1 и соответствующим документам по стандартизации (например, TR 36.814-900 и M.2135-1). Для целей проверки производительности также реализован сценарий калибровки по 3GPP

TR 36.814-900, табл. A-2.1 и проведены соответствующие тесты.

Далее будем полагать, что все точки доступа WiFi и относящиеся к ним клиенты (т. н. «блуждающие устройства») выполнены по одной и той же технологии IEEE 802.11-2012. Для того чтобы моделировать реальные сетевые структуры, такие устройства должны располагаться вокруг соответствующих точек доступа. Точки доступа могут размещаться повсеместно: в кафе, на транспорте, в магазинах, в торговых центрах и т. д. Расстояние от устрой-

Таблица 1

Основные параметры имитационного моделирования

Параметр	Значение/источник
Основные параметры	
Предел мощности устройства при конкуренции	23 дБм на интерфейс
Модель мобильности	Случайное направление движения, 3 км/ч
Период наблюдения	10 с
Система LTE	
Модель затухания	ITU-R M.2135-1 [19], табл. A.2.2-1, A1-3
Модель замираний	ITU-R M.2135-1 [19], разд. 1.3.1.1
Доступ к среде	Кольцевое расписание
Управление мощностью и скоростью передачи	SINR, целевое значение в 15 дБ
Частотные ресурсы	10 + 10 МГц FDD в каждом секторе, короткий CP
Режим сигнализации	2 из 20 специальных подкадров, кадр 10 мс
Радиооборудование	ITU-R M.2135-1 [19], табл. 8-4
Конфигурация антенны	1x2 (разнесенный прием на базовой станции eNB)
Система WiFi	
Модель затухания	Эмпирическая, основана на [20]
Модель замираний	Только корреляция, основана на [21]
Доступ к среде	CSMA/CA, мягкий порог -76 дБм
Управление мощностью и скоростью передачи	SINR, целевое значение 25 дБ
Частотные ресурсы	20 МГц TDMA
Режим сигнализации	Отдельный, контрольная скорость 18 Мб/с, RTS/CTS
Радиооборудование	Порог шума -95 дБм
Конфигурация антенны	1x1 (простая антенна)

ства до его точки доступа ограничено максимально допустимыми потерями в канале. Предположим при этом, что точки доступа и устройства не изменяют своего положения во время моделирования.

Рассмотрим случай, когда число блуждающих устройств равно максимально возможному числу активных D2D-соединений, поэтому такие устройства могут задействовать около половины имеющихся системных ресурсов в условиях максимальной выгрузки трафика. Проводимое исследование предполагает, что все WiFi-соединения используют один и тот же частотный диапазон и должны уступить любой активной передаче данных, для которой мощность превышает установленный порог. Для получения дополнительной информации о конфигурации сетей WiFi приведем табл. 1, основанную в том числе на документации драйвера с открытым исходным кодом Atheros.

Результаты проведенного моделирования.

Для того чтобы получить наиболее полную информацию о преимуществах от выгрузки сотового трафика на соединения D2D (WiFi Direct), проанализируем производительность WiFi Direct с сетевой поддержкой при различных уровнях интерференции. Не будем отдавать предпочтение моделированию какого-либо конкретного вида трафика, а вместо этого рассмотрим

различную плотность клиентских устройств, чтобы понять, как выгрузка трафика на WiFi Direct функционирует при различной загрузке. Для этого будем изменять процент WiFi Direct соединений от нуля до 30 %. Основываясь на текущей статистике трафика и поведении клиентов, будем считать маловероятным, чтобы более 30 % из них находились в пределах досягаемости D2D других клиентов.

Результаты для общей пропускной способности представлены на рис. 4. В частности, приведены уровни пропускной способности LTE и WiFi Direct в сумме на соту, основываясь на ассоциации источников данных с сотами. Видно, что выгрузка трафика LTE на линии D2D приводит к значительному увеличению пропускной способности соты, фактически удваивая ее при 30 % уровне выгрузки. Однако, если присутствуют блуждающие устройства, прирост пропускной способности более скромный, но он все еще составляет почти 50 % при том же уровне выгрузки трафика в 30 %.

Энергетическая эффективность обычно измеряется в битах на Джоуль и, следовательно, существенно зависит от свойств конкретной радиотехнологии. Поскольку энергетические показатели устройств, как правило, специфичны для их поставщиков, используем коэффициенты из табл. 2, которые не основаны на какой-либо конкретной

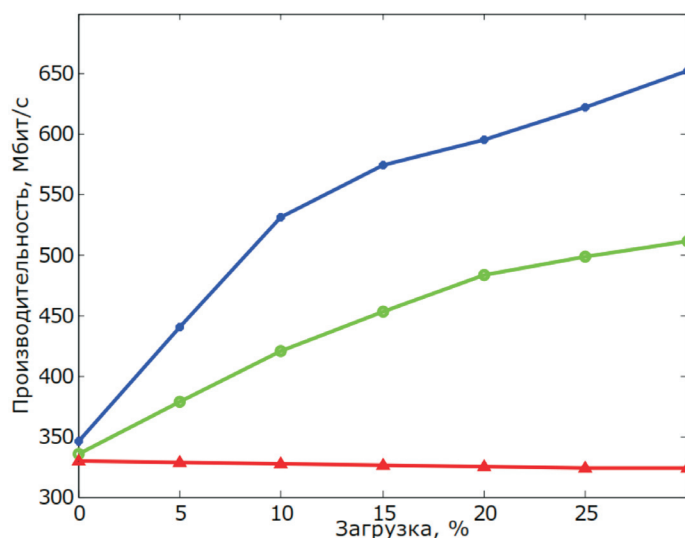


Рис. 4. Общая пропускная способность (сумма по LTE и WiFi)
 (—●—) D2D, без блуждающих устройств; (—●—) D2D, с блуждающими устройствами; (—▲—) LTE

Таблица 2

Нормированные затраты энергии

Режим функционирования	Система LTE	Система WiFi
Ожидание	0	0
Состояние покоя	0,1	0,1
Прослушивание	—	0,25
Прием данных	0,5	0,5
Передача данных	1,0	1,0

реализации. Зависимости для соответствующих значений энергетической эффективности, приведенные на рис. 5, ясно показывают, что использование WiFi значительно эффективнее по сравнению с LTE.

В частности, энергетическая эффективность WiFi более чем в четыре раза превосходит LTE в рамках рассматриваемой модели энергопотребления. Это во многом связано с более высокой скоростью передачи данных по WiFi. Кроме того, пользователям систем длительной эволюции LTE выделяются небольшие частотные фрагменты в нескольких временных слотах, поэтому их приемопередатчики должны оставаться активными в течение длительных интервалов времени, тогда как фактическая скорость передачи данных относительно мала. Для сравнения, протокол MAC WiFi активирует передатчик только тогда, когда он факти-

чески обращается к каналу и каждый раз использует всю полосу пропускания.

Даже когда у пользователей WiFi возникают задержки доступа к каналу из-за сообщений RTS или CTS, их устройства могут погрузиться в режим сна в течение этих периодов времени. В итоге только несколько радиоинтерфейсов WiFi в сети полностью задействованы в любой момент времени, и все они либо передают, либо получают данные. Результаты моделирования показывают, что время доступа к среде (т. е. время, которое пакет проводит в буфере уровня MAC и ниже) в отсутствие блуждающих устройств иногда оказывается даже меньше, чем у LTE (рис. 6). Это связано, прежде всего, с тем, что в LTE скорость передачи данных значительно ниже.

Когда блуждающие устройства присутствуют, ситуация более выгодна для LTE,

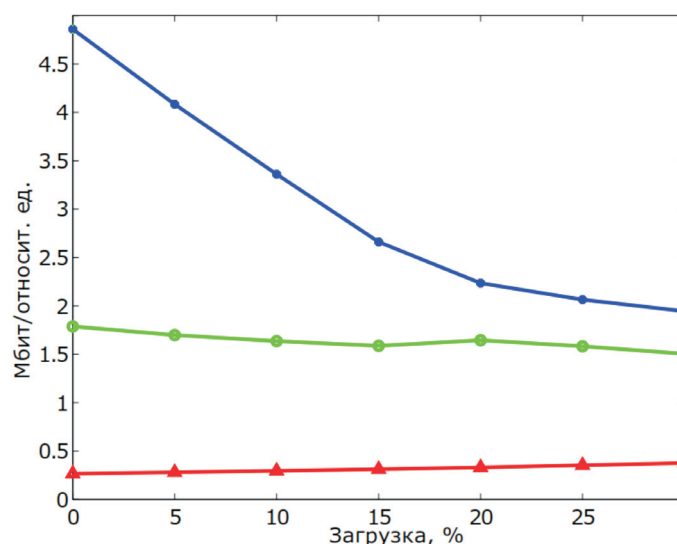


Рис. 5. Энергетическая эффективность
(—●—) D2D, без блуждающих устройств; (—●—) D2D, с блуждающими устройствами; (—▲—) LTE

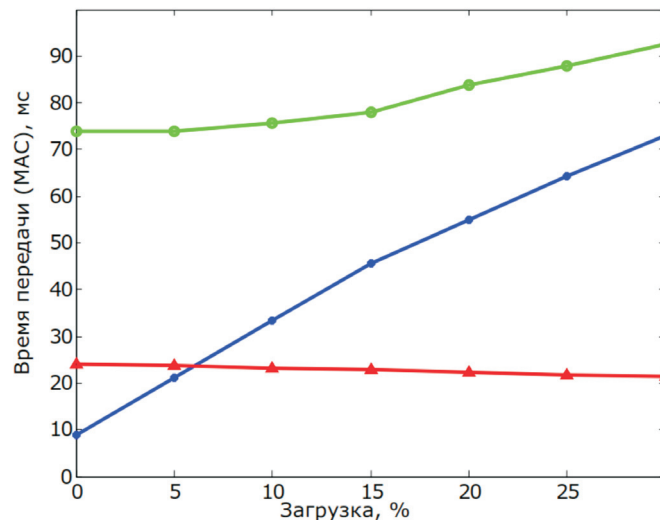


Рис. 6. Среднее время доступа к среде
(—●—) D2D, без блуждающих устройств; (—●—) D2D, с блуждающими устройствами; (—▲—) LTE

однако значительного ухудшения работы не наблюдается, как и прежде с пропускной способностью, поскольку выгружается все больше и больше трафика. Таким образом, в условиях активной интерференции выигрыш от использования связи D2D менее заметен, но все равно полностью оправдывает выгрузку трафика.

Внедрение и испытания связи D2D на модельной сети

В этом разделе рассмотрим некоторые основные особенности разработки, которая делает предложенную систему выгрузки трафика на соединения D2D с сетевой поддержкой достаточно реалистичной для интеграции с сотовыми сетями и Интернетом. На сегодняшний день ключевые аспекты протокола и сервисная архитектура уже являются частью версий 12 и 13 стандарта 3GPP LTE. В данной связи разработан прототип технологического решения для выгрузки трафика при поддержке сотовой сети, включающий в себя три основных компонента:

- 1) обслуживание на клиентской стороне, которое функционирует в фоновом режиме на пользовательских устройствах;
- 2) базу данных контента, поддерживающую хранение контента и управление доступом к нему;
- 3) сервер D2D, который также выполняет функции ProSe, управляет соединениями

и координирует их.

База данных контента представляет собой веб-платформу, предоставляющую услуги по обмену данными своим пользователям. Примером служат социальные сети, такие как Facebook, Google+ и YouTube. Сервер ProSe предназначен для оператора сотовой сети и/или поставщика услуг Интернет (ISP). Следует отметить, что альтернативно предложенной архитектуре аналогичная услуга может быть предоставлена и когда оба объекта управляются оператором. Однако при этом перенос пользователей из известных социальных сетей на новую платформу может быть более сложной задачей. Предложенное решение показывает, что база данных контента и сервер ProSe легко интегрируются в бесшовной форме в существующую инфраструктуру веб-обслуживания и не требуют каких-либо решений, которые не вписывались бы в традиционную веб-парадигму.

В испытаниях на модельной сети использовались телефоны Sony Xperia ZL с Android OS, а также Samsung Galaxy S4. Одной из основных функций, необходимых для пользовательского оборудования, была поддержка сотовых и WiFi-соединений одновременно. Из-за ограничений в потреблении энергии большинство систем избегают такой функции, поэтому необходимо было обойти систему управления

услугами на базе Android WiFi и напрямую взаимодействовать с драйвером WiFi. Аналогично GNU Linux Android предоставляет необходимые инструменты: интерфейс *wpa_supplicant*, управляемый с помощью утилиты *wpa_cli*. К сожалению, имеющиеся прошивки не позволяют получить доступ к этим утилитам даже для разработчиков, и поэтому использовалась модифицированная прошивка Cyanogenmod.

В качестве модельной сети использовалась экспериментальная сотовая система, развернутая департаментом телекоммуникаций Технологического университета

г. Брно (BUT), Чехия. LTE-система BUT является полным коммерчески пригодным решением для всех важнейших подсистем, входящих в современные сети 4G (рис. 7). Модельная сеть была сконфигурирована для обеспечения необходимых услуг, таких как VoIP, поверх конвергентной инфраструктуры радиодоступа LTE и WiFi. Ядро сети EPC обеспечивает высокую скорость передачи данных с соответствующим уровнем QoS, имея возможность поддерживать при этом до ста тысяч одновременно обслуживаемых пользователей. Для видео и голосовых услуг используется высокопро-

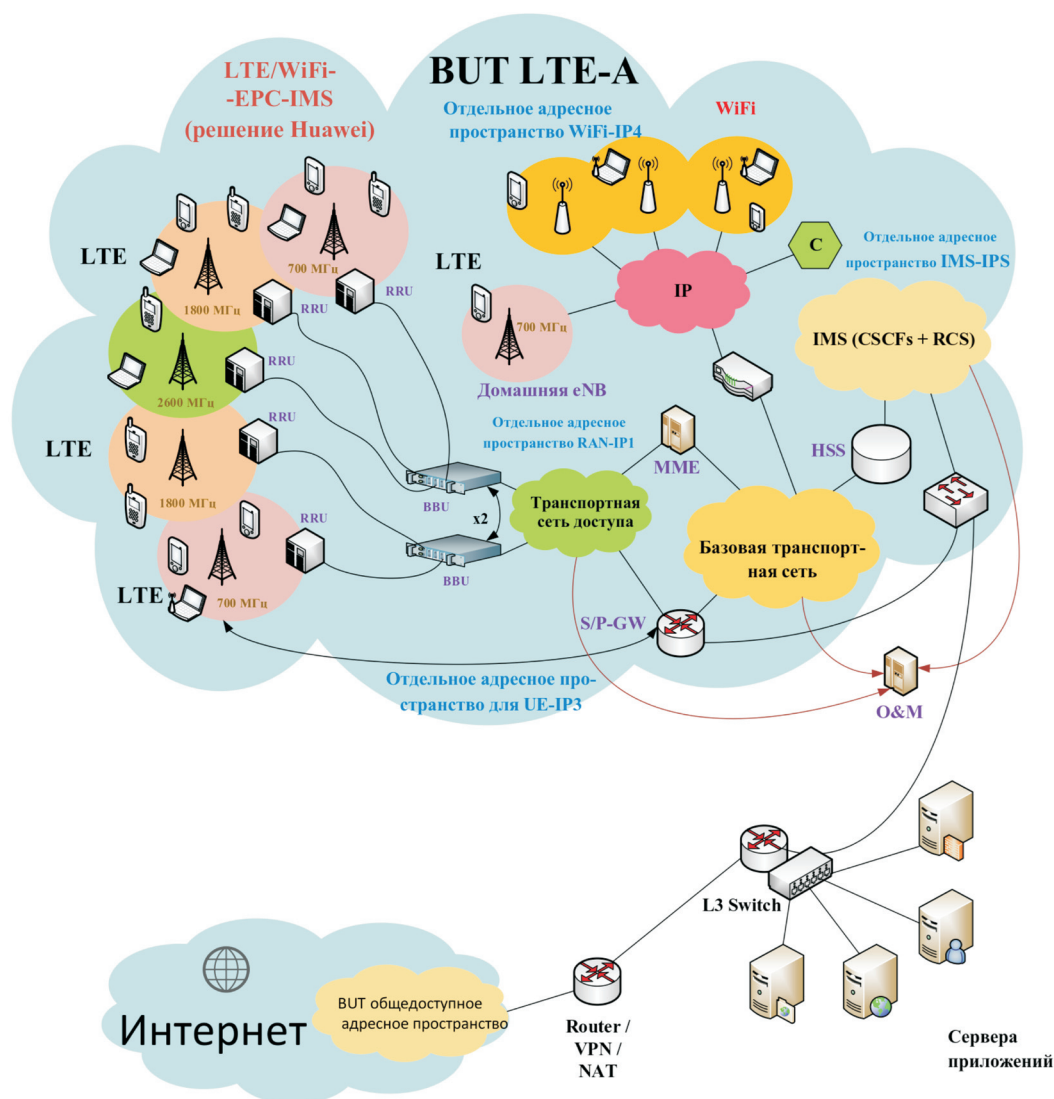


Рис. 7. Сеть LTE Технологического университета г. Брно

изводительная IP-мультимедийная подсистема (IMS).

Задачей испытаний стало доказательство наличия преимуществ при совместном использовании мобильной сети и системы D2D. Основные цели оценки производительности были сформулированы так, чтобы дополнить результаты, полученные выше при имитационном моделировании. Наибольший интерес вызывали те ситуации, когда возникали локальные перегрузки, что потенциально могло бы уменьшить эффективность работы D2D-технологий в будущих реальных сетях. Установление соответствующих пределов производительности и ограничений для технологии D2D и определение того, какие услуги могут быть наиболее подходящими для такого взаимодействия в обозримом будущем, стали важнейшими результатами проведенных испытаний.

Один из главных параметров при установлении соединения D2D – длительность интервала времени занятия линии. Этот параметр имеет решающее значение, т. к. его более низкое значение существенно снижает нагрузку на инфраструктуру сети LTE, перемещая трафик на альтернативную линию D2D, а также улучшает другие параметры работы. Время отклика при этом может быть разбито на несколько компонентов, как это описано ниже. Из-за небольшого размера сообщения другие требования к QoS для сигнализации D2D тривиальны и являются такими же, как и для любой другой сети доступа. Иные аспекты производительности линии D2D, которые напрямую

не связаны с инфраструктурой мобильной сети, выходят за рамки данной работы.

Изучая предложенный протокол для взаимодействия D2D, можно установить, что ключевым показателем является время отклика сервера, которое существенно зависит от нагрузки на соту. Полученные экспериментальные результаты представлены на рис. 8. Видно, что задержка на самом деле не достигает предельных значений, если сота не загружена почти на 100 %, но когда это происходит, вариация задержки растет до недопустимых значений. На практике система будет функционировать неудовлетворительно, если задержка превысит значение приблизительно в 100 мс. Это означает, что сигнальный трафик D2D должен получить приоритет на уровне планировщика LTE, если это возможно на практике.

Принимая во внимание то, что также может существовать некоторая дополнительная задержка, своевременное установление соединения D2D даст пользователям уверенность в получении контента вовремя, в том числе и в случаях, когда они перемещаются друг относительно друга. Можно сделать вывод о том, что в зависимости от качества сотовой связи полезность соединений D2D может отличаться. Однако даже с использованием современных технологий LTE задержки передачи информации могут быть до десяти раз больше, чем при использовании D2D поверх WiFi. Это означает, что технология D2D с использованием WiFi Direct является чрезвычайно конкурентоспособным решением.

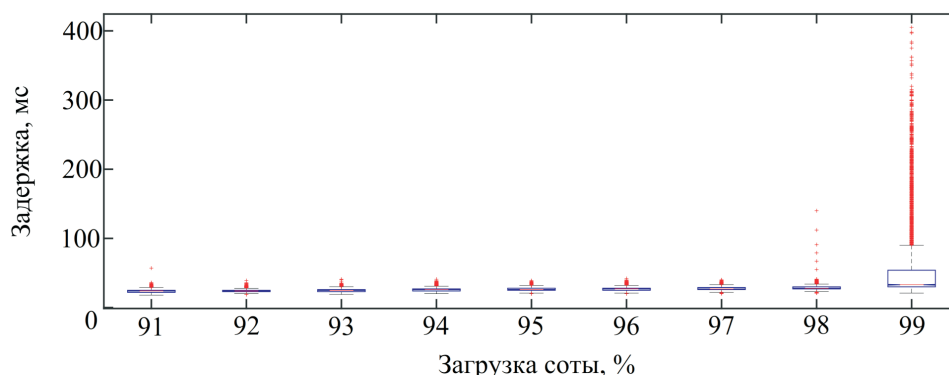


Рис. 8. Измерение задержки сервера в системе D2D

Заключение

Для операторов мобильной связи взаимодействие через D2D может стать весьма эффективным механизмом для выгрузки трафика из сотовой сети, а также для обеспечения поддержки социальных приложений посредством локализации предоставления услуг. В статье рассмотрена новая технология D2D с поддержкой сети, проведено ее исследование и описана реализация соответствующего прототипа, а также определены основные проблемы и потенциальные преимущества прямого взаимодействия между мобильными устройствами. Подчеркнем наиболее важные выводы и уроки, извлеченные из проведенных исследований.

- Успешная интеграция D2D и обеспечение совместимости с существующей архитектурой 3GPP LTE показывает, что нет никаких технических проблем, которые препятствовали бы провайдерам предлагать приложения D2D своим клиентам уже сейчас. Более того, некоторые из них могли бы существенно улучшить свою инфраструктуру и предоставлять новые услуги, которые были бы невозможны без необходимости непрерывного отслеживания местоположения с помощью ГЛОНАСС/GPS.

- Системы создания, управления и обслуживания D2D-соединений можно будет достаточно просто реализовать в рамках

разработанных протоколов. Только некоторые платформы, которые не реализуют стандартизованные стеки, могут столкнуться с определенными трудностями при управлении трафиком.

- Качество восприятия услуги для D2D-соединений будет превосходить лучшие сотовые соединения в пределах разумных расстояний между пользователями (до 50 м). При этом технология WiFi Direct позволяет передавать потоковое видео высокой четкости, а также поддерживать приложения реального времени с высокой энергетической эффективностью.

- Успех всей концепции D2D во многом зависит от поддержки мобильным оператором позиционирования в соте, а также от готовности различных операторов к взаимодействию. Это, как обычно, зависит от наличия соответствующих стандартов 3GPP и рыночных механизмов.

Все проблемы, выявленные в ходе исследования системы D2D и реализации прототипа выгрузки трафика, можно оперативно решить в ходе внедрения в сетях операторов мобильной связи. Поскольку требования пользователей к новым услугам все возрастают, можно утверждать, что предложенная в данной статье система D2D с сетевой поддержкой найдет в той или иной форме широкое применение в сетях связи при практическом внедрении мобильных систем четвертого и, в особенности, пятого поколения.

СПИСОК ЛИТЕРАТУРЫ

1. Kliks A., Holland O., Basaure A., Matimikko M. Spectrum and license flexibility for 5G networks // IEEE Communications Magazine. 2015. Vol. 53. No. 7. Pp. 42–49.
2. Asadi A., Wang Q., Mancuso V. A survey on Device-to-Device communication in cellular networks // IEEE Communications Surveys and Tutorials. 2014. Vol. 16. No. 4. Pp. 1801–1819.
3. ElSawy H., Hossain E., Alouini M.-S. Analytical modeling of mode selection and power control for underlay D2D communication in cellular networks // IEEE Transactions on Communications. 2014. Vol. 62. No. 11. Pp. 4147–4161.
4. Lee N., Lin X., Andrews J.G., Heath R.W. Power control for D2D underlaid cellular networks: modeling, algorithms, and analysis // IEEE Journal on Selected Areas in Communications. 2015. Vol. 33. No. 1. Pp. 1–13.
5. Huang Y., Nasir A.A., Durrani S., Zhou X. Mode selection, resource allocation, and power control for D2D-enabled two-tier cellular network // IEEE Transactions on Communications. 2016. Vol. 64. No. 8. Pp. 3534–3547.
6. Yu C.-H., Tirkkonen O., Doppler K., Ribeiro C. Power optimization of Device-to-Device communication underlaying cellular communication // Internat. Conf. on Communications. IEEE. 2009. Pp. 1–5.
7. Yu C.-H., Tirkkonen O. Device-to-Device underlay cellular network based on rate splitting // Wireless Communications and Networking Conf. IEEE. 2012. Pp. 262–266.

8. Yu C.-H., Doppler K., Ribeiro C., Tirkkonen O. Resource sharing optimization for Device-to-Device communication underlaying cellular networks // *IEEE Transactions on Wireless Communications*. 2011. Vol. 10. No. 8. Pp. 2752–2763.
9. Seppala J., Koskela T., Chen T., Hakola S. Network controlled Device-to-Device (D2D) and cluster multicast concept for LTE and LTE-A networks // *Wireless Communications and Networking Conf. IEEE*. 2011. Pp. 986–991.
10. Min H., Seo W., Lee J. Reliability improvement using receive mode selection in the Device-to-Device uplink period underlaying cellular networks // *IEEE Transactions on Wireless Communications*. 2011. Vol. 10. No. 2. Pp. 413–418.
11. Chen T., Charbit G., Hakola S. Time hopping for Device-to-Device communication in LTE cellular system // *Wireless Communications and Networking Conf. IEEE*. 2010.
12. Xu Y., Yin R., Han T., Yu G. Dynamic resource allocation for Device-to-Device communication underlaying cellular networks // *International Journal of Communication Systems*. 2012. No. 38.
13. Al-Kanj L., Dawy Z., Yaacoub E. Energy-aware cooperative content distribution over wireless networks: Design alternatives and implementation aspects // *IEEE Communications Surveys & Tutorials*. 2013. Vol. 15. No. 4. Pp. 1736–1760.
14. Sankaran C.B. Data offloading techniques in 3GPP Rel-10 networks: A tutorial // *IEEE Communications Magazine*. 2012. Vol. 50. No. 6.
15. Xing B., Seada K., Venkatasubramanian N. An experimental study on Wi-Fi Ad-hoc mode for mobile Device-to-Device video delivery // *INFOCOM Workshops. IEEE*, 2009. Pp. 1–6.
16. Lei L., Zhong Z., Lin C., Shen X. Operator controlled Device-to-Device communications in LTE-advanced networks // *IEEE Wireless Communications*. 2012. Vol. 19. No. 3.
17. Evolved universal terrestrial radio access (E-UTRA) // 3GPP Technical Report (TR) 36.814-900. 2010.
18. Zhuang J., Jalloul L., Novak R., Park J. IEEE 802.16m Evaluation Methodology Document (EMD). 2009.
19. ITU-R M.2135. Guidelines for evaluation of radio interface technologies for IMT-Advanced. ITU, 2009.
20. Konstantinou K., Kang S., Tzaras C. A measurement-based model for Mobile-to-Mobile UMTS Links // *Vehicular Technology Conference*. 2007.
21. Gudmundson M. Correlation model for shadow fading in mobile radio systems // *Electronics Letters*. 1991. Vol. 27. No. 23. Pp. 2145–2146.

Статья поступила в редакцию 11.06.2018.

REFERENCES

1. Kliks A., Holland O., Basaure A., Matinmikko M. Spectrum and license flexibility for 5G networks. *IEEE Communications Magazine*, 2015, Vol. 53, No. 7, Pp. 42–49.
2. Asadi A., Wang Q., Mancuso V. A survey on Device-to-Device communication in cellular networks. *IEEE Communications Surveys and Tutorials*, 2014, Vol. 16, No. 4, Pp. 1801–1819.
3. ElSawy H., Hossain E., Alouini M.-S. Analytical modeling of mode selection and power control for underlay D2D communication in cellular networks. *IEEE Transactions on Communications*, 2014, Vol. 62, No. 11, Pp. 4147–4161.
4. Lee N., Lin X., Andrews J.G., Heath R.W. Power control for D2D underlaid cellular networks: modeling, algorithms, and analysis. *IEEE Journal on Selected Areas in Communications*, 2015, Vol. 33, No. 1, Pp. 1–13.
5. Huang Y., Nasir A.A., Durrani S., Zhou X. Mode selection, resource allocation, and power control for D2D-enabled two-tier cellular network. *IEEE Transactions on Communications*, 2016, Vol. 64, No. 8, Pp. 3534–3547.
6. Yu C.-H., Tirkkonen O., Doppler K., Ribeiro C. Power optimization of Device-to-Device communication underlaying cellular communication. *International Conference on Communications*, IEEE, 2009, Pp. 1–5.
7. Yu C.-H., Tirkkonen O. Device-to-Device underlay cellular network based on rate splitting. *Wireless Communications and Networking Conference (WCNC)*, IEEE, 2012, Pp. 262–266.
8. Yu C.-H., Doppler K., Ribeiro C., Tirkkonen O. Resource sharing optimization for Device-to-Device communication underlaying cellular networks. *IEEE Transactions on Wireless communications*, 2011, Vol. 10, No. 8, Pp. 2752–2763.
9. Seppala J., Koskela T., Chen T., Hakola S. Network controlled Device-to-Device (D2D) and cluster multicast concept for LTE and LTE-A networks. *Wireless Communications and Networking Conference (WCNC)*, IEEE, 2011, Pp. 986–991.
10. Min H., Seo W., Lee J. Reliability improvement using receive mode selection in the Device-to-Device uplink period underlaying cellular networks. *IEEE Transactions on*

Wireless Communications, 2011, Vol. 10, No. 2, Pp. 413–418.

11. **Chen T., Charbit G., Hakola S.** Time hopping for Device-to-Device communication in LTE cellular system. *Wireless communications and Networking Conference (WCNC)*, IEEE, 2010.

12. **Xu Y., Yin R., Han T., Yu G.** Dynamic resource allocation for Device-to-Device communication underlying cellular networks. *International Journal of Communication Systems*, 2012, No. 38.

13. **Al-Kanj L., Dawy Z., Yaacoub E.** Energy-aware cooperative content distribution over wireless networks: Design alternatives and implementation aspects. *IEEE Communications Surveys & Tutorials*, 2013, Vol. 15, No. 4, Pp. 1736–1760.

14. **Sankaran C.B.** Data offloading techniques in 3GPP Rel-10 networks: A tutorial. *IEEE Communications Magazine*, 2012, Vol. 50, No. 6.

15. **Xing B., Seada K., Venkatasubramanian N.** An experimental study on Wi-Fi Ad-hoc mode for

mobile Device-to-Device video delivery. *INFOCOM Workshops*, IEEE, 2009, Pp. 1–6.

16. **Lei L., Zhong Z., Lin C., Shen X.** Operator controlled Device-to-Device communications in LTE-advanced networks. *IEEE Wireless Communications*, 2012, Vol. 19, No. 3.

17. *Evolved Universal Terrestrial Radio Access (E-UTRA). 3GPP Technical Report (TR) 36.814-900*, 2010.

18. **Zhuang J., Jalloul L., Novak R., Park J.** *IEEE 802.16m Evaluation Methodology Document (EMD)*, 2009.

19. *ITU-R M.2135. Guidelines for evaluation of radio interface technologies for IMT-Advanced*. ITU, 2009.

20. **Konstantinou K., Kang S., Tzaras C.** A measurement-based model for Mobile-to-Mobile UMTS links. *Vehicular Technology Conference*, 2007.

21. **Gudmundson M.** Correlation model for shadow fading in mobile radio systems. *Electronics Letters*, 1991, Vol. 27, No. 23, Pp. 2145–2146.

Received 11.06.2018.

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

ПЯТТАЕВ Александр Владиславович

PYATTAEV Alexander V.

E-mail: alex.pyattaev@gmail.com

АНДРЕЕВ Сергей Дмитриевич

ANDREEV Sergey D.

E-mail: serge.andreev@gmail.com

КУЧЕРЯВЫЙ Евгений Андреевич

KOUCHERYAVY Yevgeni A.

E-mail: ykoucheryavy@hse.ru



Программное обеспечение вычислительных, телекоммуникационных и управляющих систем

DOI: 10.18721/JCSTCS.11203

УДК 004.434

РЕЛЯЦИОННОЕ ПРОГРАММИРОВАНИЕ С МЕМОИЗАЦИЕЙ И ОТРИЦАНИЕМ

Е.А. Моисеенко, А.В. Подкопаев

Санкт-Петербургский государственный университет,
Санкт-Петербург, Российская Федерация

Реляционная парадигма позволяет описывать программы как набор отношений. Отношения не делают различий между входными и выходными параметрами. Благодаря этому одна и та же реляционная программа может использоваться для решения нескольких связанных проблем. В статье представлена реализация двух расширений реляционного программирования, полезных при разработке интерпретаторов: табличной мемоизации и конструктивного отрицания. Мемоизация позволяет эффективно обходить пространство состояний интерпретатора, а отрицание — проверять, что заданное состояние интерпретатора недостижимо. Полезность данных расширений продемонстрирована на примере интерпретатора для многопоточного императивного языка программирования.

Ключевые слова: реляционное программирование, декларативное программирование, логическое программирование в ограничениях, табличная мемоизация, конструктивное отрицание.

Ссылка при цитировании: Моисеенко Е.А., Подкопаев А.В. Реляционное программирование с мемоизацией и отрицанием // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2018. Т. 11. № 2. С. 35–46. DOI: 10.18721/JCSTCS.11203

RELATIONAL PROGRAMMING WITH MEMOIZATION AND NEGATION

E.A. Moiseenko, A.V. Podkopaev

St. Petersburg State University, St. Petersburg, Russian Federation

The relational paradigm allows to express programs as relations. Unlike functions, the relations do not distinguish input and output parameters, thus a single relational program can be used to solve several related problems. Relational interpreters are of particular interest. These interpreters can execute a program, check that the program satisfies a set of constraints or generate a program that has specified properties. In order to take advantages of the relational interpreter, the developer needs to define the semantics of the programming language as a relation. In this work, we present an implementation of two useful extensions of relational programming: tabling and constructive negation. Tabling helps to traverse the state space of the interpreter

efficiently. Constructive negation allows to check that some state of the interpreter is unreachable. We show how this extension can be used on an example of a relational interpreter for a concurrent imperative programming language.

Keywords: relational programming, declarative programming, constraint logic programming, tabling, constructive negation.

Citation: Moiseenko E.A., Podkopaev A.V. Relational programming with memoization and negation. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems, 2018, Vol. 11, No. 2, Pp. 35–46. DOI: 10.18721/JCSTCS.11203

Введение

Реляционное программирование — это вид декларативного программирования, в рамках которого программы представляются как набор отношений. Среди реляционных программ особый интерес представляют реляционные интерпретаторы, способные не только исполнять программу, но и проверять удовлетворяет ли программа набору ограничений или генерировать программу, обладающую заданными свойствами [1]. Таким образом, на базе одного реляционного интерпретатора могут быть разработаны различные языковые инструменты, такие как символьный интерпретатор, верификатор, генератор кода и др. Для того чтобы воспользоваться всеми преимуществами реляционного интерпретатора, разработчику необходимо описать семантику языка в терминах отношений.

Для задания семантики императивных программ часто используются системы помеченных переходов (СПП) [2]. В наиболее общем виде СПП — это множество состояний и отношение перехода между парой состояний по заданной метке. При реализации семантики императивного языка как СПП особенно полезными оказываются такие расширения реляционного программирования, как табличная мемоизация [9, 11] и конструктивное отрицание [3, 7, 8, 10]. Мемоизация позволяет запоминать и неоднократно использовать результаты запросов к отношениям, тем самым избегая повторных вычислений. Представив пространство состояний СПП как отношение и выполнив его мемоизацию, можно затем эффективно выполнять его обход. Конструктивное отрицание позволяет построить логическое отрицание запроса. При помощи конструк-

тивного отрицания становится возможно, например, проверить, что заданное состояние СПП недостижимо.

В рамках данной работы мы реализовали табличную мемоизацию и конструктивное отрицание для языка реляционного программирования OCanren [17]. Используя эти расширения, мы разработали интерпретатор минималистичного императивного многопоточного языка программирования, а затем применили данный интерпретатор для верификации нескольких многопоточных алгоритмов, таких как алгоритм взаимного исключения Деккера [16], алгоритм взаимного исключения Петерсона [18], алгоритм барьера и др.

1. Реляционное программирование

Парадигма реляционного программирования использует многие идеи из *логического программирования в ограничениях* (constraint logic programming) [4] и, в частности, от языков семейства Prolog [9, 11]. Одним из представителей реляционной парадигмы является язык программирования MiniKanren [1]. MiniKanren состоит из небольшого числа примитивов. Благодаря этому он может быть легко реализован как встраиваемый предметно-ориентированный язык (embedded domain specific language). Одной из таких реализаций является OCanren [18], встроенный в функциональный язык OCaml.

Основы. Процесс программирования на реляционном языке можно условно разделить на два этапа: определение типов данных и отношений и формирование запросов к отношениям. Проведя аналогию с языком SQL, можно сказать, что отношения соответствуют таблицам, а запросы — SELECT-запросам к базе данных. При

формировании запроса в реляционном языке разработчик передает некоторые аргументы отношения, а другие аргументы заменяет свободными переменными. Если отношение может быть выполнено в контексте переданных аргументов, то запрос завершается успехом (*success*) и возвращает потенциально бесконечный список подстановок для свободных переменных, на которых выполняется отношение. Если отношение не может быть выполнено, то запрос завершается неудачей (*fail*).

```
let append xs ys =
  match xs with
  | [] → ys
  | x::xs' → x::(append xs' ys)

let append° xs ys zs =
  ((xs ≡ [] ) ∧ (ys ≡ zs))
  ∨
  fresh (x xs' zs')
    (xs ≡ x::xs') ∧
    (zs ≡ x::zs') ∧
    (append° xs' ys zs')
```

Листинг 1. Функция `append` и отношение `append°`

Рассмотрим функциональную программу конкатенации списков `append` и ее реляционный аналог `appende` (листинг 1). При помощи отношения `append°` мы можем выполнить конкатенацию списков, сгенерировать все разбиения списка на пару списков или проверить, что конкатенация пустого списка с любым другим не изменяет последний (листинг 2). В ответе на последний запрос можно видеть *свободную логическую переменную*, обозначенную как `_0`. Свободная логическая переменная в ответе на запрос означает, что на ее место может быть подставлен любой терм соответствующего типа.

Рассмотрим более подробно определение отношения `append°`. Можно видеть, что определение состоит из бинарного отношения $\equiv$, вызова других отношений (в данном случае рекурсивного вызова `append°`), логических связок $\wedge$ (конъюнкции) и $\vee$ (дизъюнкции), а также конструкции `fresh`.

```
run * (append° [1] [2] q) ~> success {
  q = [1; 2]
}

run * (append° q r [1; 2]) ~> success {
  q = [],      r = [1; 2];
  q = [1],     r = [2];
  q = [1; 2], r = [];
}

run * (append° [] q q) ~> success {
  q = _0
}
```

Листинг 2. Примеры запросов к отношению `append°`

Бинарное отношение эквивалентности $t \equiv u$ определено на множестве логических термов. Два терма t и u находятся в отношении $t \equiv u$, если они унифицируются [15], т. е. существует подстановка, заменяющая вхождение свободных переменных в t и u на другие логические термы таким образом, что t и u становятся синтаксически эквивалентными. Например, термы `Cons(_0, Nil)` и `Cons(1, _1)` унифицируются, т. к. существует подстановка, которая ставит в соответствие первой переменной константу 1, а второй переменной — терм `Nil`.

Конструкция `fresh` служит для введения новых «свежих» переменных. Свежая переменная может быть унифицирована с любым другим термом. Далее в программе такая переменная может быть заменена на данный терм. В коде отношения `append°` с помощью `fresh` вводятся три новых переменных x , xs' , zs' . Эти переменные затем используются для того, чтобы декомпозировать исходные списки xs и zs на голову и хвост.

Отрицание. Помимо перечисленных выше конструкций в языке `OSanGen` также имеется *ограничение неэквивалентности* $\neq$ (disequality constraint). Ограничение $t \neq u$ выполняется тогда и только тогда, когда термы t и u не связаны отношением $t \equiv u$. С помощью ограничения $\neq$ может быть определено отношение `remove°` (листинг 3), которое связывает список ys со списком zs , в котором удалено первое вхождение x .


```

let removeo x ys zs =
  ((ys ≡ []) ∧ (zs ≡ []))
  ∨
  fresh (ys')
    (ys ≡ x::ys') ∧ (zs ≡ ys')
  ∨
  fresh (y ys' zs')
    (x ≠ y) ∧
    (ys ≡ y::ys') ∧
    (zs ≡ y::zs') ∧
    (removeo x ys' zs')

```

Листинг 3. Отношение remove^o

Отношение $\neq$ представляет собой очень ограниченную форму отрицания в реляционном программировании. К сожалению, этой формы отрицания нередко недостаточно для выражения многих полезных программ в реляционной форме. В частности, невозможно обобщить отношение remove^o таким образом, чтобы оно выполняло удаление первого элемента списка, удовлетворяющего произвольному предикату p^o (т. к. в третьем дизъюнкте отношения remove^o требуется проверка, что элемент списка y не удовлетворяет предикату p^o).

Ниже, в разделе 4, описан метод конструктивного отрицания [3, 7, 8, 10]. Данный метод впервые был предложен для языков семейства Prolog, однако в этой работе, насколько нам известно, впервые реализовано конструктивное отрицание для встроенного реляционного языка программирования.

Мемоизация. Это метод оптимизации программ, заключающийся в сохранении результата работы функции на конкретных аргументах и неоднократного использования этого результата при повторных вызовах функции [14].

Табличный метод мемоизации (*tabling*) позволяет выполнять мемоизацию отношений, которая способна существенно ускорить исполнение запросов к рекурсивным отношениям. Кроме того, некоторые запросы к мемоизированной версии отношения способны завершаться за конечное время, тогда как аналогичные запросы к немемоизированной версии не завершаются.

Табличный метод мемоизации изве-

стен в контексте языка программирования Prolog [9, 11]. Ниже, в разделе 3, мы представим улучшенный алгоритм табличной мемоизации для OCanren, поддерживающий отношения с ограничениями неэквивалентности.

2. Реляционный интерпретатор

В данном разделе описана общая структура реляционного интерпретатора, т. е. интерпретатора, заданного как отношение, для простого многопоточного императивного языка программирования. Также продемонстрируется как данный интерпретатор может использоваться для проверки инвариантов программ.

Реляционные СПП. Для того чтобы описать семантику императивного языка программирования в реляционной форме, используются системы помеченных переходов (СПП). В рамках данной статьи определим СПП как тройку (S, L, R) , где S — это множество состояний, L — множество меток, а $R \subseteq L \times S \times S$ — отношение перехода. СПП тривиальным образом представима как реляционная программа: множествам состояний S и меток L соответствуют типы данных `State` и `Label`, а отношение перехода R может быть определено как отношение $\text{step}^o :: \text{Label} \times \text{State} \times \text{State}$.

В качестве множества меток будем рассматривать множество пар $\text{Tid} \times \text{Action}$, где первый элемент является идентификатором потока, который выполняет действие, а второй элемент определяет само действие. В случае императивной программы, работающей с разделяемой памятью, действием может быть, например, чтение значения из разделяемой переменной или запись значения в разделяемую переменную.

Обход пространства состояний. При помощи отношения step^o возможно описать отношение достижимости reachable^o (листинг 4). Данное отношение параметризовано предикатом p^o и связывает начальное состояние t с некоторым достижимым состоянием t' , на котором выполняется p^o . Можно видеть, что отношение достижимости определено индуктивно. Любое состояние, на котором выполняется p^o , достижимо из самого себя. Кроме того, состояние

t'' достижимо из состояния t , если существует переход из t в t' и t'' достижимо из t' .

$$\begin{aligned} \text{let reachable}^\circ p^\circ t t'' = \\ ((p^\circ t) \wedge (t \equiv t'')) \vee \\ \text{fresh}(t') \\ (\text{step}^\circ t t') \wedge \\ (\text{reachable}^\circ t' t'') \end{aligned}$$

Листинг 4. Отношение достижимости

У приведенного выше определения reachable° имеется существенный недостаток: оно неэффективно обходит пространство состояний, поскольку одно и то же состояние может быть посещено несколько раз. Рассмотрим ромбовидную систему переходов, изображенную на рис. 1. Предположим, что требуется обойти все состояния, достижимые из состояния S_1 . Начав обход, запрос тут же обнаружит, что из начального состояния существует два перехода в состояния S_2 и S_3 . Далее запрос запустит два независимых подзапроса: один для обхода состояний, достижимых из состояния S_2 , и другой для обхода состояний, достижимых из S_3 . Оба подзапроса продолжат обход, обнаружив состояние S_4 и все состояния, достижимые из него, однако эти состояния будут посещены дважды.

Для того чтобы исправить этот недостаток, необходимо сохранять множество посещенных состояний. Данного эффекта можно добиться, применив к отношению

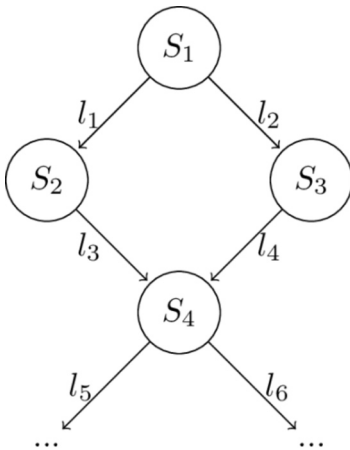


Рис. 1. Ромбовидная система переходов

reachable° табличный метод мемоизации реляционных программ. В рамках данной работы поддержка табличной мемоизации была добавлена в OSanGen.

Проверка инвариантов. В предыдущем разделе было показано, что отношение reachable° может использоваться для обхода пространства состояний системы помеченных переходов, соответствующей императивной программе. С помощью данного отношения может быть определено отношение invariant° (листинг 5), проверяющее, что для данной программы выполняется инвариант, заданный предикатом safe° . Для проверки недостижимости ошибочного состояния необходим оператор отрицания. В языке OSanGen отсутствовала поддержка отрицания, поэтому в рамках данного исследования мы разработали соответствующее расширение, основанное на методе конструктивного отрицания [3, 7, 8, 10].

$$\begin{aligned} \text{let invariant}^\circ \text{ safe}^\circ t = \neg (\\ \text{fresh}(t') \\ (\text{reachable}^\circ (\neg \text{safe}^\circ) t t') \\) \end{aligned}$$

Листинг 5. Проверка инварианта

3. Табличный метод мемоизации

Идея метода табличной мемоизации заключается в неоднократном использовании ответов на запрос. С мемоизированным отношением связывается таблица, в которой ключом является запрос (т. е. набор переданных аргументов отношения), а значением — список ответов на данный запрос. В дальнейшем, если в ходе исполнения программы вновь будет сделан запрос, эквивалентный некоторому более раннему, ответы на данный запрос будут извлечены из таблицы. Схема алгоритма табличной мемоизации представлена на рис. 2.

При выполнении запроса к отношению сначала выполняется абстракция аргументов. Абстракция аргументов позволяет уменьшить размер таблицы для хранения аргументов и списка ответов: несколько запросов с различными аргументами потенциально могут быть абстрагированы

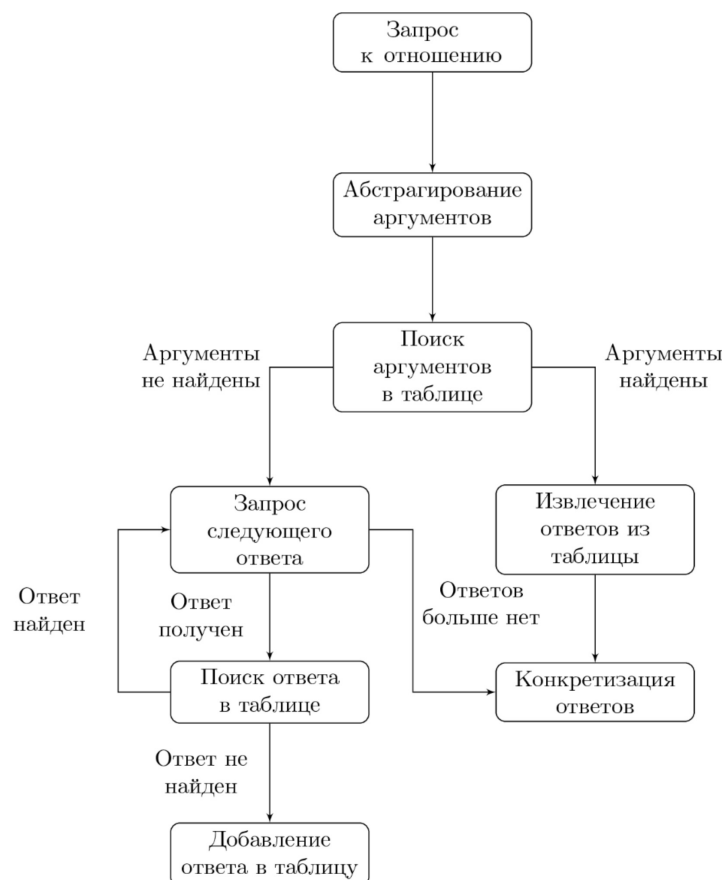


Рис. 2. Блок-схема алгоритма табличной мемоизации

до одного запроса. Следует отметить, что сильная абстракция может привести к тому, что запросы к рекурсивным отношениям перестанут завершаться, т. к. после абстракции в рекурсивный вызов попадут слишком общие аргументы. В нашей реализации при абстракции аргументов отбрасывается вся информация об ограничениях неэквивалентности. Так, например, запрос $\text{reachable}^\circ s q \{ \neq s' \}$ (где q — свободная переменная) будет абстрагирован до запроса $\text{reachable}^\circ s q$.

Полученный после абстракции список аргументов используется как ключ для поиска записи об эквивалентном запросе в таблице. Запросы считаются эквивалентными, если их аргументы *альфа-эквивалентны*, т. е. синтаксически равны с точностью до переименования переменных. В качестве структуры данных для хранения множества запросов алгоритм использует хеш-

таблицу. Перед вычислением хеш-функции от логического терма сначала выполняется переименование всех переменных в порядке обхода терма в глубину (это необходимо для того, чтобы альфа-эквивалентные термы имели одинаковое значение хеш-функции).

Если запись не была найдена в таблице, необходимо выполнить запрос к оригинальной, немемоизированной версии отношения, чтобы получить ответы. После получения очередного ответа выполняется проверка: не был ли данный ответ добавлен в таблицу ранее. Если такой ответ встречается впервые, он добавляется в таблицу.

Если запись была найдена, т. е. запрос отношения с соответствующими аргументами уже выполнялся когда-то, необходимо извлечь полученные ранее ответы из таблицы.

Перед возвращением списка ответов необходимо выполнить их конкретизацию.

Напомним, что табличный алгоритм начинается с абстракции аргументов вызова. Абстракция может удалить некоторую информацию об аргументах, чтобы привести их к более общему виду (в частности, в нашей реализации удаляется информация об ограничениях неэквивалентности). Абстрагированные аргументы затем используются для вызова оригинальной, немемоизированной версии отношения. Другими словами, полученные ответы являются ответами не на исходный запрос, а на его абстрагированную версию. Чтобы получить ответ на исходный запрос, необходимо скомбинировать информацию об аргументах, потерянную на стадии абстракции, и полученные ответы. Заметим, что возможна ситуация, при которой некоторые ответы окажутся отброшены, как не удовлетворяющие исходному запросу. Таким образом, хотя информация об ограничениях неэквивалентности отбрасывается на стадии абстракции, на этапе конкретизации происходит проверка совместимости ограничений с полученными ответами. Возвращаясь к примеру с запросом $\text{reachable}^\circ s \ q \{ \neq s' \}$, получив список ответов на его абстрагированную версию $\text{reachable}^\circ s \ q$, например, два ответа $\text{reachable}^\circ s \ s'$ и $\text{reachable}^\circ s \ s''$, алгоритм на стадии конкретизации отбросит первый ответ как не удовлетворяющий исходному запросу.

4. Конструктивное отрицание

Конструктивное отрицание [3, 7, 8, 10] — это форма отрицания в реляционном программировании, которая, получив список ответов на запрос, строит на основе этого списка набор ограничений. Немного упростив, можно сказать, что конструктивное отрицание меняет местами отношения эк-

вивалентности и ограничения неэквивалентности, а также конъюнкции и дизъюнкции.

В рамках данной работы реализовано конструктивное отрицание для языка программирования OSanGen. Эта реализация основана на теоретическом обосновании данной формы отрицания, приведенном в [8].

Стратифицированные программы. Метод конструктивного отрицания имеет ограничения и применим не ко всем реляционным программам. В частности, оператор отрицания может быть применен только к запросу, имеющему конечное количество ответов. Более того, семантика реляционной программы с отрицанием определена только для *стратифицированных* программ [8].

Для того чтобы дать определение стратифицированной реляционной программы, свяжем с каждой программой ориентированный граф, где вершинами выступают отношения. Между двумя вершинами A и B есть ребро, если в определении отношения A есть вызов отношения B . Будем помечать ребра меткой «+» если B встречается в определении A не под оператором отрицания, и меткой «-» иначе. Реляционная программа является стратифицированной, если в соответствующем графе нет циклов, содержащих хотя бы одно ребро с меткой «-».

На рис. 3 приведен пример простой стратифицированной программы, а на рис. 4 — пример нестратифицированной программы.

Граф стратифицированной программы может быть разбит на компоненты связности по отношению связности по ребрам с меткой «+». Данные компоненты называются *стратами*. Страты соединены ребрами

```
let A x y = ¬ (B x y)
let B x y = x ≡ y
```

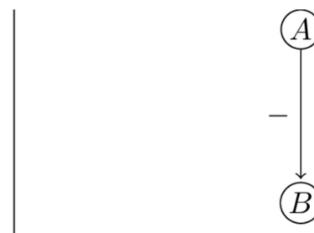


Рис. 3. Пример стратифицированной программы

```
let A x y = ¬ (B x y)
let B x y = A x y
```

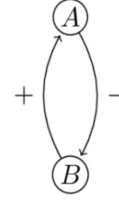


Рис. 4. Пример нестратифицированной программы

с меткой «-». Определение стратифицированной программы гарантирует, что между стратами нет циклических зависимостей.

Интуитивно, нестратифицированная программа содержит противоречащие друг другу отношения. В этом случае невозможно конструктивно построить набор ограничений при отрицании.

Семантика конструктивного отрицания. Определим процедуру конструктивного отрицания более формально. Для этого рассмотрим процедуру ответа на запрос как процесс трансляции реляционной программы в логическую формулу [8]. Напомним, что мы рассматриваем только запросы, имеющие конечное множество ответов. Для начала будем рассматривать простые программы, содержащие только отношения эквивалентности ($\equiv$). Для таких программ результатом трансляции будет логическая формула:

$$E_+ = \exists \bar{x}. \bigvee_{i=1}^N \bigwedge_{j=1}^{M_i} t_{ij} \equiv u_{ij}. \quad (1)$$

В данной формуле под кванторами существования стоят переменные, введенные в отношении с помощью конструкции *fresh*. Отношения эквивалентности $t \equiv u$ в формуле соответствуют отношениям эквивалентности в коде реляционной программы. Наконец, заметим, что конъюнкции и дизъюнкции в коде реляционной программы можно перегруппировать таким образом, чтобы получить формулу в ДНФ.

Выполнимость конъюнкции отношений эквивалентности $t \equiv u$ может быть проверена при помощи алгоритма унификации. Результатом работы данного алгоритма в случае, если набор отношений $t \equiv u$ выполним, является подстановка, связывающая свободные переменные. Список данных подстановок является от-

ветом на исходный запрос. Отрицание запроса представим в виде логической формулы:

$$\neg E_+ = \forall \bar{x}. \bigwedge_{i=1}^N \bigvee_{j=1}^{M_i} t_{ij} \neq u_{ij}. \quad (2)$$

Таким образом, список ответов на отрицание запроса должен соответствовать логическому отрицанию формулы E_+ (2). Подформулы вида $t \neq u$ напоминают ограничения неэквивалентности, однако стоит заметить, что термы t и u могут содержать универсально квантифицированные переменные из $\bar{x}$. Ограничение $\forall \bar{x}. t \neq u$ выполнимо, если не существует подстановки, связывающей только переменные из $\bar{x}$, которая унифицирует термы t и u :

$$\forall \bar{x}. t \neq u \Leftrightarrow \neg \exists \bar{x}. t \equiv u. \quad (3)$$

Таким образом, задача выполнимости ограничений обобщения также может быть решена при помощи унификации.

Теперь рассмотрим наиболее общий случай, когда логическая формула, соответствующая ответу на запрос, может содержать отношения эквивалентности и ограничения неэквивалентности с квантором всеобщности. Отрицание формулы, содержащей отношения эквивалентности и ограничения обобщения:

$$\begin{aligned} E &= \exists \bar{x}. \bigvee_{i=1}^N ((\bigwedge_{j=1}^{M_i} t_{ij} \equiv u_{ij}) \wedge \\ &\quad \wedge (\forall \bar{y}. \bigwedge_{k=1}^{L_i} \bigvee_{l=1}^{Q_k} p_{kl} \neq u_{kl})) \\ \neg E &= \forall \bar{x}. \bigwedge_{i=1}^N ((\bigvee_{j=1}^{M_i} t_{ij} \neq u_{ij}) \vee \\ &\quad \vee (\exists \bar{y}. \bigvee_{k=1}^{L_i} \bigwedge_{l=1}^{Q_k} p_{kl} \equiv u_{kl})). \end{aligned} \quad (4)$$

Формула (4) показывает, что при отрицании данной формулы получается дизъюнкция двух подформул. Левая подформула является ограничением неэквивалентности, а правая похожа на формулу положитель-

ной программы, с тем отличием, что термы p и u могут содержать как экзистенциально квантифицированные переменные из u , так и универсально квантифицированные переменные из x . Таким образом, ограничения в правом дизъюнкте имеют вид $\forall x \exists y. p \equiv r$. Выполнимость конъюнкции данных ограничений может быть проверена с помощью модифицированного алгоритма унификации [7]. Алгоритм принимает на вход множество универсально квантифицированных переменных U , множество экзистенциально квантифицированных переменных E , а также множество ограничений ($t \equiv u$), связанных конъюнкцией. Алгоритм возвращает `true` и подстановку, связывающую только свободные переменные, в случае если ограничения выполнимы, и `false` — в противном случае.

Логическая формула, соответствующая всей стратифицированной реляционной программе, строится поэтапно. Выполняется топологическая сортировка страт реляционной программы. Затем транслируется подпрограмма, соответствующая последней в порядке сортировки страте. Затем рассматривается предыдущая в порядке сортировки страта. Выполняется трансляция соответствующей ей подпрограммы, при этом все вхождения оператора отрицания заменяются на отрицание логической формулы, полученной на предыдущем этапе. После этого процесс повторяется для предыдущей страты. Так продолжается, пока не будут обработаны все страты. В итоге будет получена логическая формула, соответствующая исходному запросу. Выполнимость данной формулы может быть сведена к последовательности унификаций.

5. Апробация

Для апробации предложенных в данной работе расширений языка *OSanGen* была рассмотрена задача применения реляционного интерпретатора многопоточных программ (см. раздел 2) к проблемам верификации программ и синтеза кода синхронизации.

Перед выполнением апробации реляционный интерпретатор многопоточных программ был обобщен для возможности

исполнения программ в различных *моделях памяти*. Модель памяти определяет семантику взаимодействия потоков с разделяемой памятью. Модель *последовательной согласованности* (sequential consistency) [6] является наиболее простой. В рамках данной модели каждое возможное поведение программы является результатом некоторого чередования инструкции различных потоков. Современные языки программирования и аппаратные архитектуры предоставляют разработчикам более сложные модели, называемые *слабыми моделями памяти*. Помимо модели SC в нашем интерпретаторе реализована поддержка модели Total Store Order (TSO) [13], а также модели Strong Release-Acquire (SRA) [5].

В задачах верификации на вход интерпретатору подавалась многопоточная программа, а также инвариант, выполнимость которого следовало проверить. Были рассмотрены программа передачи сообщения между двумя потоками, алгоритм взаимного исключения Деккера [16], алгоритм взаимного исключения Коэна [12], а также алгоритм барьера для двух потоков. Была произведена попытка верификации данных алгоритмов в трех моделях памяти: SC, TSO и SRA. Известно, что алгоритм Деккера без использования дополнительных операций синхронизации корректен только в модели SC. Реляционный интерпретатор подтвердил этот факт, найдя контрпримеры в случае исполнения алгоритма Деккера в моделях TSO и SRA. Два других алгоритма были успешно верифицированы во всех трех моделях. В табл. 1 приведено среднее время работы интерпретатора для данных тестов с 95 % доверительным интервалом, полученное по десяти запускам.

В задаче синтеза синхронизации были рассмотрены те же алгоритмы, но на вход интерпретатору подавался шаблон программы, в котором были опущены *модификаторы доступа*. Модификаторы доступа определяют уровень синхронизации при чтении или записи разделяемой переменной в слабых моделях. На место модификаторов доступа были подставлены свободные переменные. Задачей реляционного интерпретатора было найти подстановку

Таблица 1

Результаты применения интерпретатора для верификации

Название теста	SC	TSO	SRA
Message Passing	0.01±0.01	0.02±0.00	0.02±0.01
Dekker Lock	0.11±0.02	1.08±0.01	0.21±0.01
Peterson Lock	0.26±0.01	2.91±0.01	0.48±0.01
Cohen Lock	0.11±0.00	0.35±0.00	0.21±0.01
Barrier	0.08±0.01	0.14±0.01	0.12±0.00

Таблица 2

Результаты применения интерпретатора для синтеза синхронизации

Название теста	TSO	SRA
Message Passing	0.05±0.00	0.36±0.01
Dekker Lock	76.49±0.51	OOM
Peterson Lock	5300.51±10.09	OOM
Cohen Lock	1.00±0.05	2.48±0.07
Barrier	2.39±0.04	238.67±1.79

для этих переменных. В модели SC все операции над разделяемыми переменными полностью упорядочены, по этой причине задача генерации модификаторов доступа в данной модели неактуальна. Результаты замеров (среднее и 95 % доверительный интервал времени работы по десяти запускам) приведены в табл. 2.

Из результатов апробации следует, что реляционный интерпретатор хорошо справляется с задачей верификации. В моделях TSO и SRA пространство состояний программ, как правило, существенно больше, чем в модели SC. По этой причине время верификации для этих моделей также увеличивается. С задачей генерации модификаторов доступа интерпретатор справляется намного хуже. Для алгоритма Деккера и Петерсона в модели SRA интерпретатор завершил работу по причине исчерпания виртуальной памяти (out of memory). В модели TSO на данных тестах интерпретатор завершил свою работу, однако это заняло достаточно много времени. Такие результаты связаны с тем, что каждый раз, когда реляционный интерпретатор встречает свободную переменную в исходной программе, он делает недетерминированный выбор. Таким

образом, пространство поиска может расти экспоненциально с ростом количества свободных переменных в программе.

6. Обзор связанных работ

Одно из приложений языка Mini-Kanren — разработка реляционных интерпретаторов. В частности, в [1] представлен реляционный интерпретатор для подмножества функционального языка программирования Racket. Показано, как данный интерпретатор может использоваться для генерации программ по набору тестов или для генерации квинтов (quines) — программ, результатом выполнения которых является исходный текст программы. В отличие от этой работы, мы представляем реляционный интерпретатор для многопоточного императивного языка программирования и демонстрируем, как он может использоваться для проверки инвариантов программ.

Табличный метод мемоизации давно известен в контексте семейства языков программирования Prolog. Например, в языке XSB Prolog [11] табличная мемоизация применяется ко всем отношениям по умолчанию.

Метод конструктивного отрицания,



впервые предложенный в [3], ранее рассматривался в контексте логических языков программирования. В [8] описаны теоретические аспекты конструктивного отрицания, а также приведен алгоритм исполнения логической программы с конструктивным отрицанием для стратифицированных программ. Работа [10] обобщает конструктивное отрицание до логических программ с ограничениями. В [7] описана реализация конструктивного отрицания для более широкого класса программ, чем стратифицированные программы. Несмотря на существование множества работ, описывающих теоретические аспекты конструктивного отрицания, насколько известно нам, все современные реализации языка Prolog используют более простой метод «отрицание как неудача» (*negation as a failure*). Данный метод некорректен (*unsound*) в случае, если в запросе под отрицанием встречаются свободные переменные. В нашей статье предпринята попытка реализовать конструктивное отрицание для встроеного реляционного языка программирования.

Заключение

В данной статье представлена реализация двух расширений реляционного языка

OSanren: табличной мемоизации и конструктивного отрицания. Табличная мемоизация позволяет разрабатывать эффективные интерпретаторы на реляционном языке программирования. Конструктивное отрицание увеличивает выразительную силу реляционного языка, позволяя выражать новые отношения.

Используя язык OSanren с данными расширениями, мы разработали реляционный интерпретатор для многопоточных императивных программ. Данный интерпретатор позволяет исследовать поведение многопоточных программ, а также проверять их инварианты.

В ходе апробации продемонстрировано, что реляционный подход может использоваться для верификации и генерации синхронизации в небольших многопоточных программах, но для применения данного подхода к более сложным программам требуются некоторые улучшения. Интеграция реляционного языка с современными решателями формул в теориях (SMT solvers) может ускорить проверку выполнимости ограничений. Другим направлением работы является применение техник суперкомпиляции и частичного исполнения.

СПИСОК ЛИТЕРАТУРЫ

1. Byrd W.E., Ballantyne M., Rosenblatt G., Might M. A unified approach to solving seven programming problems (functional pearl) // Proc. of the ACM on Programming Languages. 2017. Vol. 1. ICFP. P. 8.
2. Baier C., Katoen J.P., Larsen K.G. Principles of model checking. MIT press, 2008.
3. Chan D. Constructive negation based on the completed database // Proc. of ICLP-88. 1988.
4. Jaffar J., Maher M.J. Constraint logic programming: A survey // The Journal of Logic Programming. 1994. Vol. 19. Pp. 503–581.
5. Lahav O., Giannarakis N., Vafeiadis V. Taming release-acquire consistency // ACM SIGPLAN Notices. ACM, 2016. Vol. 51. No. 1. Pp. 649–662.
6. Lamport L. How to make a multiprocessor computer that correctly executes multiprocess programm // IEEE Transactions on Computers. 1979. No. 9. Pp. 690–691.
7. Liu J.Y., Adams L., Chen W. Constructive negation under the well-founded semantics // The Journal of Logic Programming. 1999. Vol. 38. No. 3. Pp. 295–330.
8. Przymusiński T.C. On constructive negation in logic programming. Cambridge, Massachusetts, MIT Press, 1989.
9. Schrijvers T., Demoen B., Warren D.S. TCHR: a framework for tabled CLP // Theory and Practice of Logic Programming. 2008. Vol. 8. No. 4. Pp. 491–526.
10. Stuckey P.J. Constructive negation for constraint logic programming // Proc. of 6th Annual IEEE Symposium on Logic in Computer Science. IEEE, 1991. Pp. 328–339.
11. Swift T., Warren D.S. XSB: Extending Prolog with tabled logic programming // Theory and Practice of Logic Programming. 2012. Vol. 12. No. 1-2. Pp. 157–187.
12. Turon A., Vafeiadis V., Dreyer D. GPS: Navigating weak memory with ghosts, protocols, and separation // ACM SIGPLAN Notices. ACM, 2014. Vol. 49. No. 10. Pp. 691–707.
13. Sewell P., Sarkar S., Owens S., Zappa F.N., Myreen M.O. x86-TSO: a rigorous and

usable programmer's model for x86 multiprocessors // *Communications of the ACM*. 2010. Vol. 53. No. 7. Pp. 89–97.

14. **Michie D.** Memo functions and machine learning // *Nature*. 1968. Vol. 218. No. 5136. P. 19.

15. **Baader F., Snyder W.** Unification theory // *Handbook of automated reasoning*. 2001. Vol. 1. Pp. 445–532.

Статья поступила в редакцию 09.05.2018.

REFERENCES

1. **Byrd W.E., Ballantyne M., Rosenblatt G., Might M.** A unified approach to solving seven programming problems (functional pearl). *Proceedings of the ACM on Programming Languages*, 2017, Vol. 1, ICFP, P. 8.

2. **Baier C., Katoen J.P., Larsen K.G.** *Principles of model checking*. MIT press, 2008.

3. **Chan D.** Constructive negation based on the completed database. *Proceedings of ICLP-88*, 1988.

4. **Jaffar J., Maher M.J.** Constraint logic programming: A survey. *The Journal of Logic Programming*, 1994, Vol. 19, Pp. 503–581.

5. **Lahav O., Giannarakis N., Vafeiadis V.** Taming release-acquire consistency. *ACM SIGPLAN Notices*. ACM, 2016, Vol. 51, No. 1, Pp. 649–662.

6. **Lamport L.** How to make a multiprocessor computer that correctly executes multiprocess program. *IEEE Transactions on Computers*, 1979, No. 9, Pp. 690–691.

7. **Liu J.Y., Adams L., Chen W.** Constructive negation under the well-founded semantics. *The Journal of Logic Programming*, 1999, Vol. 38, No. 3, Pp. 295–330.

8. **Przymusiński T.C.** *On constructive negation in logic programming*. Cambridge, Massachusetts, MIT Press, 1989.

9. **Schrijvers T., Demoen B., Warren D.S.** TCHR: a Framework for Tabled CLP. *Theory and Practice of Logic Programming*, 2008, Vol. 8, No. 4., Pp. 491–526.

16. **Dijkstra E.W.** Cooperating sequential processes // *The Origin of Concurrent Programming*. New York: Springer, 1968. Pp. 65–138.

17. **Kosarev D., Boulytchev D.** Typed embedding of a relational language in OCaml // *International Workshop on ML*. 2016.

18. **Peterson G.L.** Myths about the mutual exclusion problem // *Information Processing Letters*. 1981. No. 12(3). Pp. 115–116.

10. **Stuckey P.J.** Constructive negation for constraint logic programming. *Proceedings of 6th Annual IEEE Symposium on Logic in Computer Science*, IEEE, 1991. Pp. 328–339.

11. **Swift T., Warren D.S.** XSB: Extending Prolog with tabled logic programming. *Theory and Practice of Logic Programming*, 2012, Vol. 12. No. 1-2, Pp. 157–187.

12. **Turon A., Vafeiadis V., Dreyer D.** GPS: Navigating weak memory with ghosts, protocols, and separation. *ACM SIGPLAN Notices*. ACM, 2014, Vol. 49, No. 10, Pp. 691–707.

13. **Sewell P., Sarkar S., Owens S., Zappa F.N., Myreen M.O.** x86-TSO: a rigorous and usable programmer's model for x86 multiprocessors. *Communications of the ACM*, 2010, Vol. 53, No. 7, Pp. 89–97.

14. **Michie D.** Memo functions and machine learning. *Nature*, 1968, Vol. 218, No. 5136, P. 19.

15. **Baader F., Snyder W.** Unification theory. *Handbook of automated reasoning*, 2001, Vol. 1, Pp. 445–532.

16. **Dijkstra E.W.** Cooperating sequential processes. *The Origin of Concurrent Programming*. Springer, New York, NY, 1968, Pp. 65–138.

17. **Kosarev D., Boulytchev D.** Typed embedding of a relational language in OCaml. *International Workshop on ML*, 2016.

18. **Peterson G.L.** Myths about the mutual exclusion problem. *Information Processing Letters*, 1981, No. 12(3), Pp. 115–116.

Received 09.05.2018.

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

МОЙСЕЕНКО Евгений Александрович

MOISEENKO Evgenii A.

E-mail: e.moiseenko@2012.spbu.ru

ПОДКОПАЕВ Антон Викторович

PODKOPAEV Anton V.

E-mail: a.podkopaev@2009.spbu.ru

DOI: 10.18721/JCSTCS.11204

УДК 681.5

МОДЕЛИРОВАНИЕ ВОЗДУХОРАЗДЕЛИТЕЛЬНОЙ УСТАНОВКИ НИЗКОГО ДАВЛЕНИЯ КАК ОБЪЕКТА УПРАВЛЕНИЯ

И.И. Суляев¹, Д.В. Седов²

¹ Норильский государственный индустриальный институт,
г. Норильск, Российская Федерация;

² «Норильский никель», г. Норильск, Российская Федерация

Рассмотрена задача разработки математической модели процесса разделения воздуха как объекта управления. Разделение воздуха — многостадийный процесс, включающий в себя несколько подготовительных стадий. Взаимодействие между стадиями (подсистемами) имеет сложный характер, а недоступность измерения внутренних переменных не дает возможности построения отдельных моделей подсистем многоуровневой системы.

Представлена концептуальная модель ручного управления установкой, в которой выделено несколько уровней принятия решений. Проведен системный анализ объекта, выбран технологический режим, для которого предложена структура линейной динамической модели. Согласно принципу последовательного раскрытия неопределенности рассчитаны параметры модели. Разработана схема активной идентификации модели на действующей установке. Представлены результаты активной и пассивной идентификации в режиме нормальной эксплуатации. Проведен анализ результатов экспериментов и причины расхождений, обусловленных в основном конструктивными особенностями установки.

Ключевые слова: воздухоразделение, кислород, идентификация, математическая модель, системный анализ.

Ссылка при цитировании: Суляев И.И., Седов Д.В. Моделирование воздухоразделительной установки низкого давления как объекта управления // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2018. Т. 11. № 2. С. 47–53. DOI: 10.18721/JCSTCS.11204

MODELING OF A LOW-PRESSURE AIR SEPARATOR AS A CONTROL OBJECT

I.I. Sulyaev¹, D.V. Sedov²

¹ Norilsk Industrial Institute, Norilsk, Russian Federation;

² Norilsk Nickel, Norilsk, Russian Federation

The article deals with the task of developing a mathematical model of an air separator as the control object. We have carried out systematic analyses of the object

and built a conceptual model with the principle of revelation consistency according to vagueness. The results of active and passive identification are presented.

The analysis of the experiment results is carried out. The causes of divergences, which are determined mainly because of structural peculiarities, are established.

Keywords: air-separation, oxygen, identification, mathematical model, systematic analysis.

Citation: Sulyaev I.I., Sedov D.V. Modeling of a low-pressure air separator as a control object. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems, 2018, Vol. 11, No. 2, Pp. 47–53. DOI: 10.18721/JCSTCS.11204

Введение

Производство цветных металлов активно использует кислород для интенсификации процессов горения во внутреннем объеме плавильных печей (печь Ванюкова, печь взвешенной плавки и др.). Технологический кислород производится в воздухо-разделительных установках (ВРУ) низкого давления методом глубокого охлаждения [2, 3, 6, 8].

На рис. 1 представлена концептуальная модель системы ручного управления ВРУ.

Лицо, принимающее решение (ЛПР) первого уровня, передает ЛПР второго уровня значения производственного задания G_r , которое в соответствии с текущими значениями расхода Y_o и чистоты продуктов разделения Y_k формирует воздействие G_o на управляющее устройство УУ_о подсистемы

объема О. В условиях влияния неконтролируемых возмущений внешней среды F ЛПР первого уровня формирует управляющее воздействие согласно производительности ВРУ и его технического состояния. Отметим, что от объемного расхода продуктов зависит и их качество. Исполнительная подсистема ИМ_о, формирующая расход воздуха на воздухо-разделительную установку в зависимости от количества задействованных компрессоров, отнесена к объекту управления.

В условиях влияния неконтролируемых возмущений (температура, степень загрязненности и уровень влажности окружающего воздуха) ручное управление не обеспечивает требуемую точность поддержания заданного объемного расхода воздуха и качества конечных продуктов. В работах [8, 9]

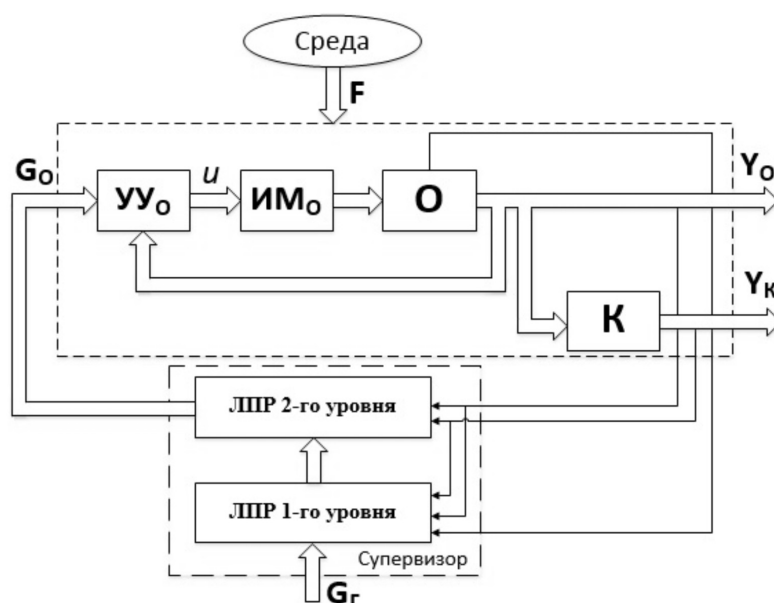


Рис. 1. Концептуальная модель системы управления ВРУ

рассмотрены различные подходы к автоматизации отдельных узлов ВРУ низкого давления. Дальнейшие исследования по автоматизации управления установками по разделению воздуха на азот, технический и технологический кислород требуют построения соответствующих математических моделей.

Разработка линейной стационарной детерминированной модели ВРУ

Системный анализ и декомпозиция процесса разделения воздуха выявляет многоуровневую структуру ВРУ (рис. 2). Нижний уровень иерархии ($L = 0$) представляют агрегаты, первый уровень ($L = 1$) образуют подсистемы сжатия, охлаждения, очистки, конденсации и разделения. Воздухоразделительная установка в целом образует уровень $L = 2$.

Модель ВРУ, в принципе, может быть получена агрегированием моделей подсистем. Известны достоинства моделей с раскрытой структурой причинно-следственных взаимосвязей подсистем. Анализ свойств объектов и систем по таким моделям дает выводы как констатирующего, так и объяснительного характера [1, 4, 7]. В случае ВРУ не удастся моделировать подсистемы аналитическим способом, а недоступность измерения внутренних переменных

установки исключает возможность экспериментального способа (идентификации). Сказанное заставляет строить модель ВРУ в целом как объект управления с закрытой внутренней структурой.

Топология модели ВРУ представлена на рис. 3. При анализе априорных данных (экспертная оценка, техническая и научная литература, ретроспективные данные) определены переменные процесса: u – расход воздуха; y_1 – расход азота; y_2 , y_3 – расход технического и технологического кислорода; f_1 , f_2 – температура воздуха на входе установки и температура охлаждающей воды [9, 10]. Основные допущения и ограничения адекватности модели: постоянство уровня кубовой жидкости в ректификационных колоннах; постоянные обороты турбодетандера; малые отклонения от разгрузочного режима. Принята гипотеза о независимости каналов.

Согласно принципу последовательного раскрытия неопределенности [4, 7], следующим этапом построения модели является определение структур операторов. Преобразователем сигнала по каждому каналу выступает устойчивое звено первого порядка с запаздыванием.

Схема пассивной идентификации параметров модели ВРУ:

анализ данных за год (интервалов рабо-

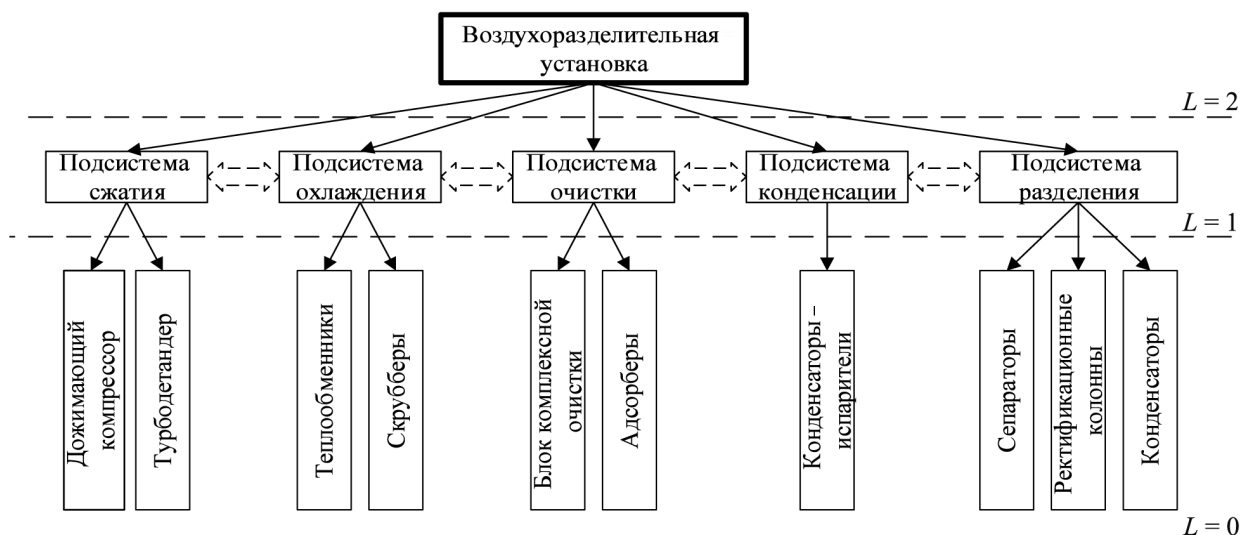


Рис. 2. Иерархическая структура воздухоразделительной установки низкого давления

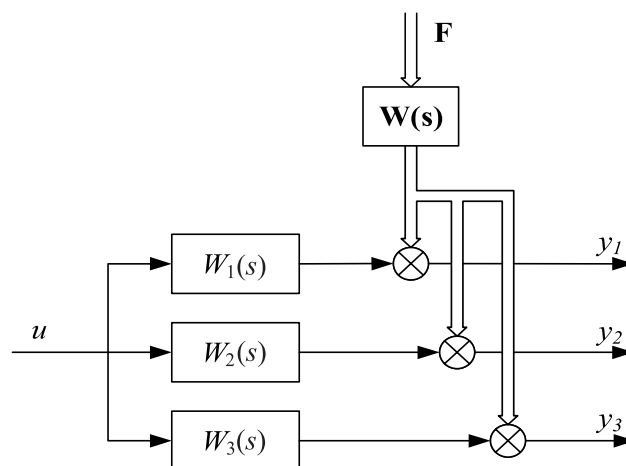


Рис. 3. Топология математической модели установки разделения воздуха

ты на разгрузочном режиме, информации о значимых факторах процесса);

фильтрация шумов и исключение экстремальных значений;

расчет параметров передаточных функций по каждому каналу;

величина запаздывания принимается на основе мнения экспертов (не менее 5 мин).

Результаты пассивной идентификации приведены в табл. 1.

Схема активной идентификации модели ВРУ:

эксперимент проводится опытным персоналом в режиме нормальной эксплуатации ВРУ (разгрузочный режим);

входная переменная u формируется как импульсное управляющее воздействие (последовательное подключение компрессорных установок в общий коллектор) с малым отклонением от начального значения (не более 10 %);

длительность управляющего воздействия формируется на основе представления экспертов об инерционности процесса (не менее 5 мин);

Таблица 1

Результаты пассивного эксперимента

Номер канала	Характер процесса	Параметры передаточной функции		
		k	T	τ
1	Устойчивый	0,15	140	300
2	Устойчивый	0,05	140	300
3	Устойчивый	0,15	140	300

Таблица 2

Таблица активной идентификации

	Базовые значения				Амплитуда импульса, %	Длительность импульса, мин	Характер процесса		
	u	y_1	y_2	y_3			y_1	y_2	y_3
1	169510	25357	8459	25349	+5	15	Устойчивый		
2	169950	25456	8489	25497	-5	15	Устойчивый		

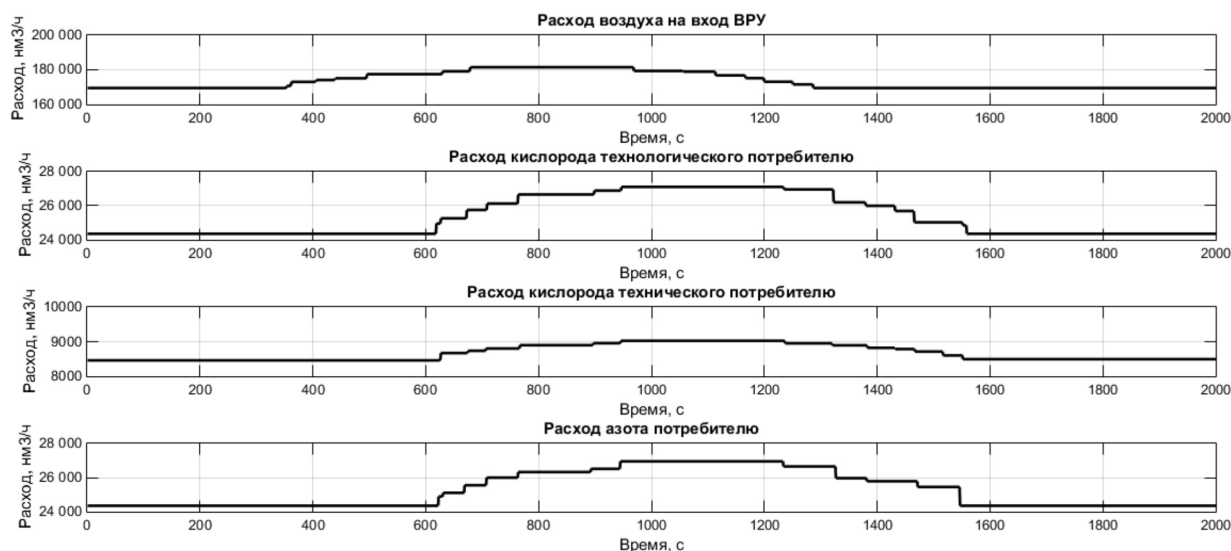


Рис. 4. Характер изменения переменных процесса при активном эксперименте

возмущения f_1 и f_2 в процессе эксперимента постоянны.

При выполнении активного эксперимента была сформирована табл. 2, которая отражает длительность импульса и характер процесса по каналам. Отметим, что по всем каналам процесс обладает самовыравниванием, что подтверждает выводы из пассивных экспериментов.

На рис. 4 представлены исторические тренды изменения входных и выходных переменных процесса при первом эксперименте. Время запаздывания составляет 5 мин, что обусловлено протяженностью пневмосистемы ВРУ. График имеет ступенчатую форму по причине переключения регенераторов четырех воздухоразделительных установок, одновременно работающих от общего воздушного коллектора.

Характер изменения выходных переменных по каналу управления соответ-

ствует структуре устойчивого звена первого порядка. Параметры модели отражены в табл. 3. Подстройка параметров модели осуществлялась средствами Matlab/System Identification Toolbox по отфильтрованным значениям переменных (среднее скользящее с окном 120 значений) с использованием алгоритма подстройки по критерию максимального правдоподобия [5].

Анализ результатов пассивного и активного экспериментов (табл. 1, 2) показывает расхождение значений постоянной времени, что объясняется переменным количеством и производительностью одновременно работающих компрессоров. Увеличение времени переходного процесса при пассивном эксперименте связано с влиянием неконтролируемых возмущений (длительный запуск компрессора из резерва, ограничения по мощности компрессора, несогласованность действий персонала).

Таблица 3

Результаты активной идентификации

Номер канала	Характер процесса	Параметры передаточной функции		
		k	T	τ
1	Устойчивый	0,14576	58,741	275
2	Устойчивый	0,0495	62,78	275
3	Устойчивый	0,14872	64,16	275

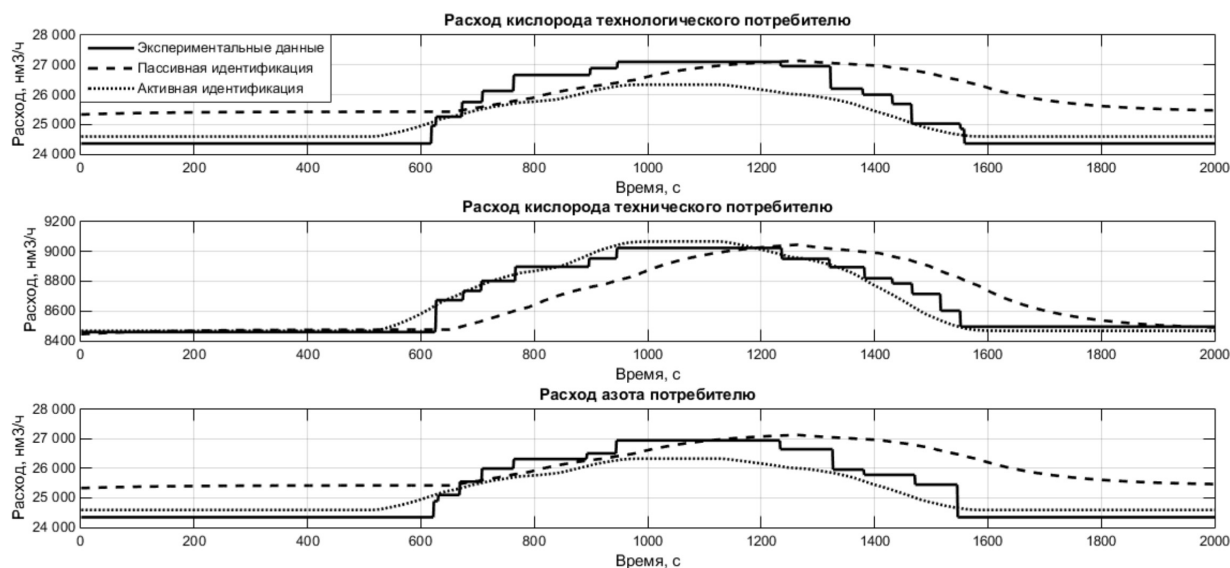


Рис. 5. Сравнение экспериментальных и теоретических данных

Таблица 4

Сравнение результатов активного и пассивного экспериментов

Номер канала	КК		СКО	
	Активная идентификация	Пассивная идентификация	Активная идентификация	Пассивная идентификация
1	0,98	0,74	6,34	7,1
2	0,98	0,76	4,22	5,17
3	0,97	0,75	6,78	7,22

Результаты моделирования представлены на рис. 5. Для проверки степени адекватности математической модели рассчитаны коэффициент корреляции (КК) и среднеквадратическое отклонение (СКО) по всем каналам управления (табл. 4). Активный эксперимент позволил увеличить КК на 32 % и уменьшить СКО на 11 %.

Заключение

В статье проведен системный анализ ВРУ низкого давления как объекта управления. Показан сложный характер взаимодействия

технологических узлов и подсистем установки. Выделены переменные технологического процесса, характеризующие объемный расход компонентов разделения воздуха. Предложена структура математической модели в классе ЛТИ-систем. Проведена пассивная и активная идентификация на действующей ВРУ низкого давления по разработанным схемам в режиме нормальной эксплуатации (разгрузочный режим). Модель может послужить основой для разработки системы управления и тренажера ВРУ с целью обучения технологического персонала.

СПИСОК ЛИТЕРАТУРЫ

1. Алексеев А.А., Кораблев Ю.А., Шестопалов М.Ю. Идентификация и диагностика систем. М.: Изд. центр «Академия», 2009. 352 с.
2. Бродянский В.М., Меерзон Ф.И. Про-

изводство кислорода. М.: Metallurgy, 1970. 2-е изд. 384 с.

3. Давыдов Н.И. Станции технологического кислорода. М.: Metallurgy, 1964. 352 с.

4. Душин С.Е., Зотов Н.С., Имаев Д.Х. Теория автоматического управления. М.: Высш. школа, 2009. 568 с.

5. Дьяконов В.П. MATLAB. Анализ, идентификация и моделирование систем: Спец. справ. М.: СОЛОН-Пресс, 2003. 576 с.

6. Епифанова В.И. Разделение воздуха методом глубокого охлаждения // Технология и оборудование: в 2 т. М.: Машиностроение, 1973.

7. Льюнг Л. Идентификация систем. Теория для пользователя. М.: Наука, 1991. 432 с.

8. Свит Т.Ф. Основы разделения воздуха ме-

тодом глубокого охлаждения. Барнаул: Изд-во АлтГТУ, 2005. 132 с.

9. Седов Д.В. Централизованное управление процессом разделения воздуха // Сб. статей межд. конф. «Таймырские чтения – 2016». Норильск: НГИИ, 2016. С. 210–213.

10. Суляев И.И. Концептуальная модель иерархической системы управления процессом плавления медно-никелевого сырья // Сб. докладов конф. «ИТЦ-2014». СПб.: ОАО «Концерн ЦНИИ «Электроприбор», 2014. С. 211–220.

Статья поступила в редакцию 18.05.2017.

REFERENCES

1. Alekseyev A.A., Korablev Yu.A., Shestopalov M.Yu. *Identifikatsiya i diagnostika sistem* [Identification and diagnostics of systems]. Moscow: Publishing Center "Academy", 2009. 352 p. (rus)

2. Brodyanskiy V.M., Meyerzon F.I. *Proizvodstvo kisloroda* [Production of oxygen]. Moscow: Metallurgiya Publ., 1970, 384 p. (rus)

3. Davydov N.I. *Stantsii tekhnologicheskogo kisloroda* [Technological oxygen stations]. Moscow: Metallurgiya Publ., 1964, 352 p. (rus)

4. Dushin S.Ye., Zotov N.S., Imayev D.Kh. *Teoriya avtomaticheskogo upravleniya* [Theory of automatic control]. Moscow: Vysshaya shkola Publ., 2009, 568 p. (rus)

5. Dyakonov V.P. *MATLAB. Analiz, identifikatsiya i modelirovaniye sistem* [Analysis, identification and modeling of systems]. Moscow: SOLON-Press Publ., 2003, 576 p. (rus)

6. Yepifanova V.I. *Razdeleniye vozdukha metodom glubokogo okhlazhdeniya* [Air separation by deep cooling]. *Tekhnologiya i oborudovaniye*

[Technology and equipment]. Moscow: Mashinostroyeniye Publ., 1973. (rus)

7. Lyung L. *Identifikatsiya sistem. Teoriya dlya polzovatelya* [Identification of systems. Theory for the user]. Moscow: Nauka Publ., 1991, 432 p. (rus)

8. Svit T.F. *Osnovy razdeleniya vozdukha metodom glubokogo okhlazhdeniya* [Basics of air separation by deep cooling]. Barnaul: AltGTU Publ., 2005, 132 p. (rus)

9. Sedov D.V. *Tsentralizovannoye upravleniye protsessom razdeleniya vozdukha* [Centralized control of the air separation process]. *Proc. of the Conference «Taymyrskiy chteniye – 2016»*, Norilsk: NGII, 2016, Pp. 210–213. (rus)

10. Sulyayev I.I. *Kontseptualnaya model iyerarkhicheskoy sistemy upravleniya protsessom plavleniya medno-nikelevogo syrya* [A conceptual model of the hierarchical control system for the melting process of copper-nickel raw materials]. *Proc. of the Conference ITTs-2014*, St. Petersburg: «Elektropribor» Publ., 2014, Pp. 211–220. (rus)

Received 18.05.2017.

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

СУЛЯЕВ Ильгиз Ильгизович

SULYAEV Ilgiz I.

E-mail: ilgizfinland@mail.ru

СЕДОВ Дмитрий Владимирович

SEDOV Dmitriy V.

E-mail: dvsedov95@mail.ru



CIRCUITS
AND
SYSTEMS
SOCIETY

ISSCS 2019

14-th International Symposium on Signals, Circuits, and Systems
Iasi, Romania, July 11–12, 2019

organized by:

Faculty of Electronics, Telecommunications and Information Technology, "Gh. Asachi" Technical University of Iasi
IEEE Romania Section & CAS Chapter in cooperation with IEEE CAS Society

Call for Papers

Honorary Chair

Daniela Tărniceriu

Technical University of Iasi,
Romania

General Co-Chairs

Liviu Goras

Technical University of Iasi,
Romania

Sanjit Mitra

University of California,
Santa Barbara, CA, USA

Technical Program Committee

Majid Ahmadi, Canada - **Chair**

David Allstot, USA

Marco Carli, Italy

Mrityunjay Chakraborti, India

Jonathon Chambers, U.K.

Pak-Chung Ching, Hong Kong

Farhat Fnaiech, Tunisia

Daniel Foty, USA

Gabriel Gomes, Brazil

Jiwu Huang, China

Gordana Jovanovic-Dolecek,
Mexico

Alexander S. Korotkov, Russia

Anamitra Makur, Singapore

Rui Martins, Macao

Yoshikazu Miyanaga, Japan

Soo-Chang Pei, Taiwan

Branimir Reljin, Serbia

Markus Rupp, Austria

Corneliu Rusu, Romania

Mohamad Sawan, Canada

Erchin Serpendin, USA

Mani Soma, USA

Wonyong Sung, Korea

Ioan Tabus, Finland

Siddik Yarman, Turkey

ISSCS 2019 aims at bringing together scientists and researchers from academia and industry to present and discuss some of their latest ideas and results.

Prospective authors are invited to submit original papers in the following areas:

- **Linear and Nonlinear Circuits & Systems**
- **Circuits & Systems for Communications**
- **Power Electronic Circuits**
- **Analog and Digital Signal Processing**
- **Image and Video Processing**
- **VLSI Circuits & Systems**

The official language of the symposium will be English.

AUTHOR'S SCHEDULE:

Proposals for special sessions	December 17, 2018
Submission of full paper	February 18, 2019
Notification of acceptance	April 15, 2019
Submission of camera-ready version and registration form with signed copyright form	May 1, 2019

The papers should be written in IEEE/CAS A4 format (double column including paper title, authors' names and affiliation, short abstract). They should not exceed 4 pages (including figures and references).

Authors are expected to present their papers at the Symposium. At least one author of each accepted paper must register in order to have the paper included in the program and Proceedings.

ISSCS 2019

<http://scs.etti.tuiasi.ro/isscs2019>

Faculty of Electronics & Telecommunications
Bd. Carol I no.11A
700506, Iasi, ROMANIA
Conference email address:

isscs2019@etti.tuiasi.ro

DOI: 10.18721/JCSTCS.11205

UDC 004

SEQUENCE-TO-SEQUENCE BASED ENGLISH-CHINESE TRANSLATION MODEL

Tian Zhaolin, Zhang Weiwei

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

In recent years, with the continuous improvement of theory in artificial intelligence, artificial neural networks has become novel tools for machine translation. Compared with traditional Statistical Machine Translation (SMT), neural network based Neural Machine Translation (NMT) transcends SMT in many aspects such as translation accuracy, long distance reordering, syntax, tolerance to noisy data et al. In 2014, with the emergence of sequence-to-sequence (seq2seq) models and attentional mechanisms introduced into the model, NMT was further refined and its performance was getting better and better. This article uses the current popular sequence-to-sequence model to construct a neural machine translation model from English to Chinese. In addition, this paper uses Long-Short Term Memory (LSTM) to replace the traditional RNN in order to solve the problem of gradient disappearance and gradient explosion that it faces in long-distance dependence. The attention mechanism has also been introduced into this article. It allows neural networks to pay more attention to the relevant parts of the input sequences and less to the unrelated parts when performing prediction tasks. In the experimental part, this article uses TensorFlow to build the NMT model described in the article.

Keywords: NMT, seq2seq, LSTM, attention mechanism, encoder-decoder, tensorflow.

Citation: Tian Zhaolin, Zhang Weiwei. Sequence-to-Sequence based english-chinese translation model. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems, 2018, Vol. 11, No. 2, Pp. 55–63. DOI: 10.18721/JCSTCS.11205

МОДЕЛЬ SEQUENCE-TO-SEQUENCE В АНГЛО-КИТАЙСКОМ ПЕРЕВОДЕ

Тянь Чжаолин, Чжан Вэйвэй

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

В последние годы, в связи с постоянным совершенствованием теории искусственного интеллекта, новыми инструментами машинного перевода стали искусственные нейронные сети. Нейронный машинный перевод (NMT) имеет

значительные преимущества по сравнению с традиционно используемым методом статистического машинного перевода (SMT) в таких аспектах, как точность перевода, изменение порядка слов в длинных предложениях, синтаксис, помехоустойчивость и т. д. После того как в 2014 году появились модели перевода по схеме «последовательность-в-последовательность» (seq2seq) и механизмы внимания, введенные в модель, методы NMT продолжали совершенствоваться, улучшалась их производительность. В данной статье для построения модели нейронного машинного перевода с английского на китайский использована популярная в настоящее время схема перевода seq2seq. Кроме того, вместо традиционно применяемой рекуррентной нейронной сети в статье для решения возникающей проблемы взрыва и исчезновения градиента для длинных строк использован метод долгой краткосрочной памяти (LSTM). Рассмотрен механизм, позволяющий нейронным сетям уделять больше внимания соответствующим частям входных последовательностей и меньше — несвязанным частям при выполнении задач прогнозирования. В экспериментальной части статьи для построения описанной модели NMT использован TensorFlow.

Ключевые слова: NMT, seq2seq, LSTM, механизм внимания, кодировщик-декодер, TensorFlow.

Ссылка при цитировании: Тянь Чжаолинь, Чжан Вэйвэй. Модель Sequence-to-Sequence в англо-китайском переводе // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2018. Т. 11. № 2. С. 55–63. DOI: 10.18721/JCSTCS.11205

1. Introduction

Introducing more reforms and implementing the One Belt One Road strategy, China is increasingly participating in international affairs. However, due to the peculiarity of the Chinese language, it is difficult for non-native speakers to master Chinese in a short period of time. In addition, due to the differences in grammatical logic between Chinese and western languages such as English and Russian, traditional statistical machine translation often fails to achieve satisfactory results. As more and more Chinese people travel around the world, and people in other countries are increasingly interested in China, Chinese, the world's most spoken language, and English, the world's most widely used language, produce an inevitable intersection.

Continuous improvement of relevant theories of artificial intelligence in the field of machine translation and the continuous popularization of high-performance hardware devices in the 21st century have paved the way for large-scale application of artificial neural networks in machine translation and created a rare opportunity for further development of neural machine translation. In 2013, Kalchbrenner and Blunsom proposed an end-to-end encoder-decoder model for machine

translation. However, the traditional RNN used in the decoder has a problem of gradient disappearance and gradient explosion, making the model difficult to practically handle long-distance dependences. Also in 2013, Graves et al. applied deep bi-directional LSTM to speech recognition, paving the way for deeper applications of bi-directional LSTM in Neural Machine Translation (NMT). In 2014, Cho et al. proposed a new sequence-to-sequence model and used LSTM (actually, LSTM is a variation of RNN) instead of traditional RNN as encoder and decoder. In the same year, Bengio et al. introduced the attention mechanism into NMT so that the neural network can pay more attention to the relevant part of the input sequences and pay less attention to the unrelated part when performing prediction tasks.

This paper uses a mature seq2seq model to construct a translation model from English to Chinese. The structure of this paper is roughly as follows. The second part introduces data sources and data preprocessing. The third part introduces the theoretical part of the encoder, attention mechanism and decoder in the seq2seq model in detail. The fourth part introduces the experimental results, the approximate implementation of the model, and the evaluation results of the model. The

fifth section introduces recommendations for further research in the future.

2. Data Source and Data Preprocessing

The data used in this paper comes from the United Nations Parallel Corpus [1]. The English-Chinese parallel corpus contains almost fifteen million sentences. All these materials contain content which was produced and manually translated from 1990 to 2014, including sentence level alignments.

Before building our seq2seq model, we need to do some preprocessing with the parallel corpus.

- Handling training/testing dataset: We extract 100,000 sentences from the parallel corpus as the testing dataset and the remains are the training dataset.
- Handling source sentences: Add “BOS” in the beginning of sentences and “EOS” in the end of sentences.
- Handling dictionaries: Generate two dictionaries for Chinese and English based on the training dataset.
- Handling unknown words: If some vocabularies/words from the testing dataset do not exist in those two dictionaries, use “UNK” to replace them.
- Handling input sequences: Generate one-hot vectors based on the original sentence and two dictionaries. Then combine these one-hot vectors as input sequences.

3. The Model

Encoder. Assuming the input sequence $\mathbf{x} = (x_1, \dots, x_T)$, the traditional recurrent neural network (RNN) calculates the hidden state vector $\mathbf{h} = (h_1, \dots, h_T)$ and output $\mathbf{y} = (y_1, \dots, y_T)$ by iterating the following equations from $t = 1$ to T :

$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h), \quad (1)$$

$$y_t = W_{hy}h_t + b_y, \quad (2)$$

where W_{xh} denotes the input-hidden weight matrix, W_{hh} denotes the hidden-hidden weight matrix, W_{hy} denotes the hidden-output weight matrix, b_h denotes the hidden bias vector, b_y denotes the output bias vector and $\mathcal{H}$ is the hidden layer function.

However, we found out that the Long-

Short Term Memory [2] has its advantages by using a gate mechanism in dealing with long distance dependences. So we use the LSTM cell proposed by Gers et al. in 2002 [3]. So here in our model, $\mathcal{H}$ is implemented by the following equations:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \quad (3)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \quad (4)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \quad (5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o), \quad (6)$$

$$h_t = o_t \tanh(c_t), \quad (7)$$

where σ denotes the logistic sigmoid function, and i , f , o , c denote the input gate, the forget gate, the output gate and the cell activation vectors respectively, all of which are the same size as the hidden vector h .

A disadvantage of the traditional recurrent neural networks is that they are not able to take advantage of subsequent context. In this paper we use a bidirectional recurrent neural network [4] to process input sequences in both directions by using two separate hidden layers, and then feedforward to a same output layer. The calculation of the forward hidden state $\bar{h}$, the backward hidden state $\tilde{h}$ and the output sequence y is shown below:

$$\bar{h}_t = \mathcal{H}(W_{x\bar{h}}x_t + W_{\bar{h}\bar{h}}\bar{h}_{t-1} + b_{\bar{h}}), \quad (8)$$

$$\tilde{h}_t = \mathcal{H}(W_{x\tilde{h}}x_t + W_{\tilde{h}\tilde{h}}\tilde{h}_{t+1} + b_{\tilde{h}}), \quad (9)$$

$$y_t = W_{\bar{h}y}\bar{h}_t + W_{\tilde{h}y}\tilde{h}_t + b_y. \quad (10)$$

Here we can also use LSTM, which is mentioned above, to replace the traditional recurrent neural network cell [5, 6]. As a result, bi-directional long-short term memory is the basic structure of the model in this paper.

Furthermore, it has been proved that the performance of a deep neural network is always better than of that with a single layer. In our case, it is totally possible to stack several layers of bidirectional RNN to generate a deep bidirectional RNN [9]. Assuming all the hidden layers are sharing the same function, the calculation of the hidden state in n^{th} layer is shown below:

$$\bar{h}_t^n = \mathcal{H}(W_{\bar{h}^{n-1}\bar{h}^n} + W_{\bar{h}^n\bar{h}^n}\bar{h}_{t-1}^n + b_{\bar{h}}^n), \quad (11)$$

$$\bar{h}_t^n = \mathcal{H}(W_{\bar{h}^{n-1}\bar{h}^n} + W_{\bar{h}^n\bar{h}^n}\bar{h}_{t-1}^n + b_{\bar{h}}^n). \quad (12)$$

If we define $\bar{h}^0 = \bar{h}^0 = \mathbf{x}$, then the output of network y_t is:

$$y_t = W_{\bar{h}^ny} \bar{h}_t^n + W_{\bar{h}^ny} \bar{h}_t^n + b_y. \quad (13)$$

Attention Mechanism. In this paper, we implement a global attention mechanism [7, 8] in our model. Firstly, we take the hidden state h_t at the top layer of deep LSTM and generate a probability distribution based on the context vector c_t to help predict the current target word y_t . The equations are shown below:

$$\tilde{h}_t = \tanh(W_c[c_t; h_t]), \quad (14)$$

$$p(y_t | y < t, x) = \text{softmax}(W_s \tilde{h}_t), \quad (15)$$

where $\tilde{h}_t$ denotes attentional vector.

The idea of the global attention mechanism is to consider all hidden states of the encoder when calculating the context vector c_t . In this mechanism, by comparing the current target hidden state h_t with every source hidden state $\bar{h}_s$, we may get a variable-length alignment vector a_t , whose size equals the number of time steps on the source side:

$$a_t(s) = \frac{\exp(\text{score}(h_t, \bar{h}_s))}{\sum_s \exp(\text{score}(h_t, \bar{h}_s))}. \quad (16)$$

The score can be calculated in three ways, all of which are shown below:

$$\text{score}(h_t, \bar{h}_s) = \begin{cases} h_t^T \bar{h}_s & \text{dot} \\ h_t^T W_a \bar{h}_s & \text{general} \\ W_a[h_t; \bar{h}_s] & \text{concat} \end{cases} \quad (17)$$

In our model, we use the general score (the second one in Eq. (17)), which has been proved to be the best one [7], to compute the alignment vector a_t . Consider the alignment vector is weights, we use the weighted average over all the source hidden states to generate the context vector c_t . An example of the attentional vector is shown in Fig. 1 [8].

Decoder. The decoder is trained to predict the next word y_t based on the given context vector c and all the previously predicted words $\{y_1, \dots, y_{t-1}\}$ [10, 11]. The calculating function is shown below:

$$p(y) = \prod_{t=1}^T p(y_t | \{y_1, \dots, y_{t-1}\}, c), \quad (18)$$

where $y = (y_1, \dots, y_T)$. Here, the conditional probability in the recurrent neural network can be also defined as:

$$p(y_t | \{y_1, \dots, y_{t-1}\}, c) = g(y_{t-1}, h_t, c), \quad (19)$$

where g denotes a nonlinear function that

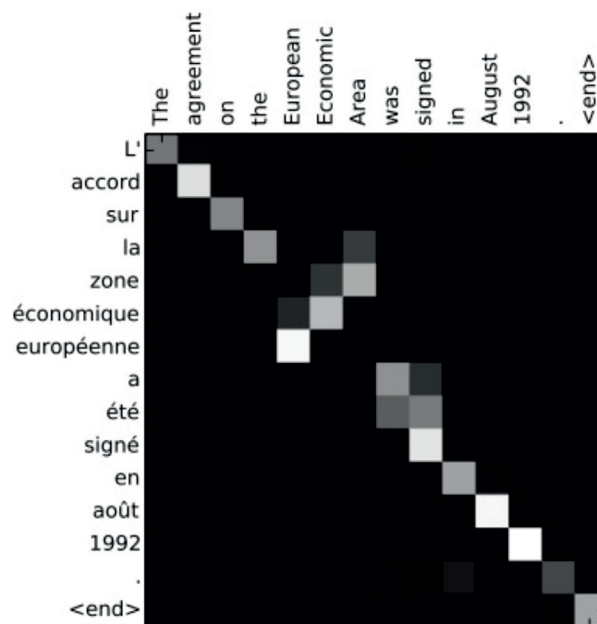


Fig. 1. English-French sample alignments found by RNN search-50. (Bahdanau et al., 2014)

outputs the probability of y_t , h_t denotes the hidden state of recurrent neural network [12].

In our model, every conditional probability described in Eq. (18) is defined as

$$p(y_t|y_1, \dots, y_{t-1}, x) = g(y_{t-1}, h_t, c_t), \quad (20)$$

where h_t is computed by the following equation:

$$h_t = f(h_{t-1}, y_{t-1}, c_t). \quad (21)$$

Be advised, the probability here is conditioned on a distinct context vector c_t for each target word y_t . The context vector c_t can be obtained by the method described in the previous section.

4. Experiments

Dependency. First our model is running on Linux. And we need the following tools to be ready.

- Python ≥ 3.5
- TensorFlow ≥ 1.2
- Numpy ≥ 1.12

It is preferable to have a GPU to help speed up the training process [25].

Model Structure. Fig. 2. shows a brief structure of our model. Here in our model, we have a double-layer bidirectional LSTM as an encoder and two-layer LSTM as a decoder. The detailed initial configuration of the encoder and the decoder is shown below.

Encoder

- Hidden state size: 1024
- Number of layers: 2
- Input keep probability: 1.0
- Output keep probability: 1.0

Decoder

- Hidden state size: 1024
- Number of layers: 2
- Input keep probability: 1.0
- Output keep probability: 1.0

Other configurations

- Learning rate: 0.0005
- Batch size: 128
- Beam size: 5
- Size of attentional vector: 512

Model Evaluation and Result Analysis. First, we use cross entropy as the loss of our model. Fig. 3 shows the variation of cross-entropy during the iteration: as the training progresses, the

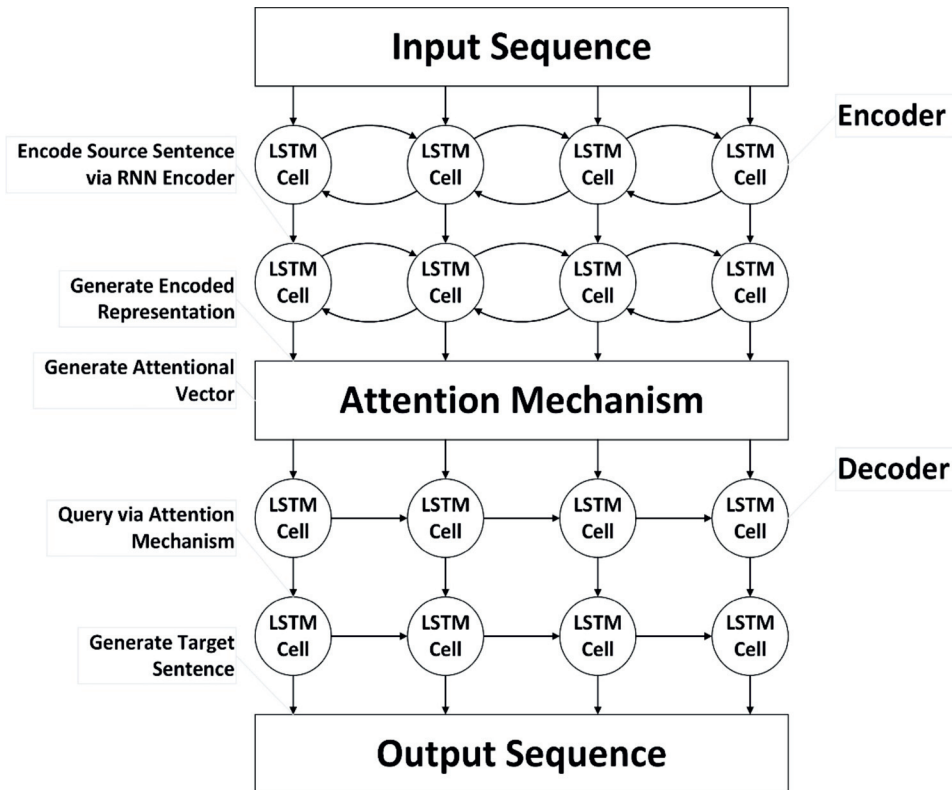


Fig. 2. The structure of the EN-CH model

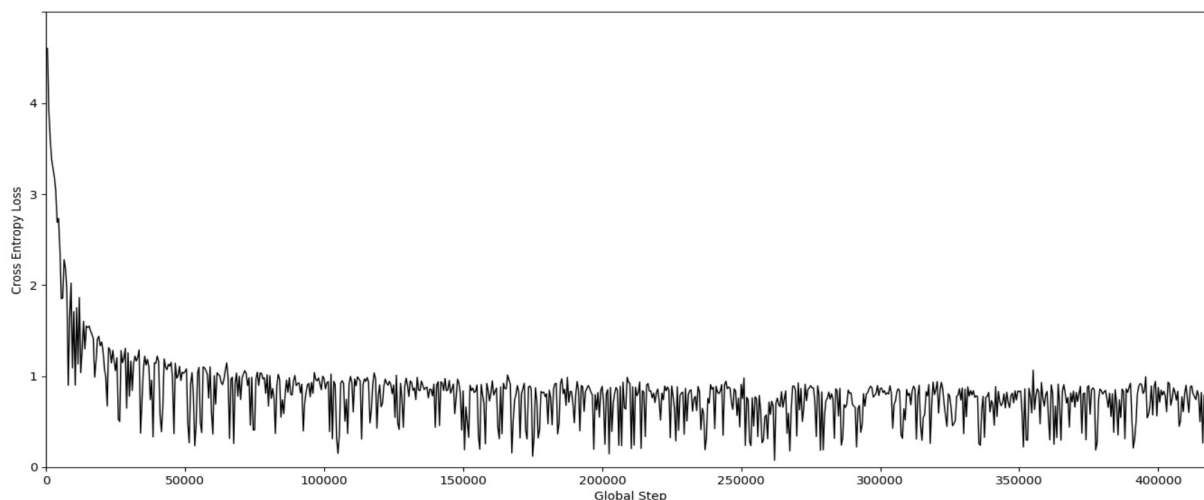


Fig. 3. Cross Entropy Loss

cross entropy decreases gradually and settles at around 0.5 and between 0.1 and 0.9, showing that the model is still quite effective.

We also notice that the losses vary in a relatively large range after the model was trained over 50,000 steps. That's because the model performs better with shorter sentences than with longer ones. When input sentences are short, the translations by the model are exactly the same with the standard translation most of the time, and, therefore, the losses in this moment may be close to zero. When input sentences are long, the translations by the model may not be accurate, but still acceptable, or, compared to

the standard translations, express the idea in another way. All of this is the reason why the losses under such circumstances are relatively larger than those for short sentences.

Table 1 shows some examples of short sentences.

Table 2 shows some examples of long sentences.

Secondly, we use a BLEU score, a kind of self-evaluation method for a machine translation model, to evaluate our model. The BLEU score is computed based on the test dataset. BLEU score is computed by the following equations [13]:

Table 1

Example of short sentences

Step/Loss	Translations by model	Standard Translations
229500/ 0.360127	纽约办事处	纽约办事处
	(议程项目5)	(议程项目5)
	联合国发展援助框架	联合国发展援助框架
277000/ 0.335871	秘书长办公厅	秘书长办公厅
	占总数的百分比	总量的百分比
	秘书处的说明	秘书处的说明
283500/0.315215	议程项目40	议程项目40
	给秘书长的信	给秘书长的信
	其他亚洲国家	亚洲其他国家



Table 2

Examples of long sentences

Step/Loss	Translations by model	Standard Translations
276500/0.715091	委员会关切地注意到， 现有住房方案没有充分解决穷人的住房需要。	委员会关切地注意到， 现行住房方案没有充分满足贫困者的住房要求。
	有人提醒说， 对某些发展中国家来说， 缓解行动可能集中在这个部门	缔约方提请注意， 对某些发展中国家来说， 缓解行动很可能会集中于这一部门。
	在同次会议上， 理事会决定将这个项目交由理事会主席协商。	在同次会议上， 理事会决定将这一项目提交理事会主席举行的磋商会议。
277500/0.894594	在生物多样性高的地区建立基于市场的可持续发展， 通过可持续利用自然资源解决农村创收发展问题；	在高度生物多样化区域， 以基于市场的方法进行可持续的社区发展， 并通过可持续地使用自然资源解决农村创收问题；
	由于外地办事处报告的数据列于财务报表附注中， 儿童基金会总部必须合理保证非消耗性财产数字准确。	由于外地办事处报告的这些数字都要在财务报表附注中披露， 儿童基金会总部必须合理地保证非消耗性财产数字的准确性。
	委员会面前的一个关键问题是发展筹资， 2001年在召开一次关于这一问题的高级别政府间会议上已经开展了筹备工作。	第二委员会所要审议的一个基本问题是为发展筹措资金； 将于2001年专门讨论这个问题的政府间高层会议正在准备之中。

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{1-r/c} & \text{if } c \leq r \end{cases} \quad (22)$$

$$Bleu = BP \cdot \exp\left(\sum_{n=1}^N \omega_n \log p_n\right), \quad (23)$$

$$\log Bleu = \min\left(1 - \frac{r}{c}, 0\right) + \sum_{n=1}^N \omega_n \log p_n, \quad (24)$$

where BP denotes the Brevity Penalty, c denotes the length of the output sentence, r denotes the length of the standard translation sentence, N equals 4 and ω_n equals 0.25.

In our model, we got a 26.6 bleu score for the English-Chinese task.

5. Further Development of NMT

In the rapidly developing and highly

competitive environment, the NMT technology is making significant progress. NMT will also be continuously improved in many aspects, including:

- Rare word problem [14, 15]
- Use of single-language data [16, 17]
- Multilingual Translation / Multilingual

NMT [18]

- Memory mechanism [19]
- Language fusion [20]
- Coverage issues [21]
- Training process [22]
- A priori knowledge fusion [23]
- Multi-modal translation [24]

Acknowledgement

We would like to thank Professor E.A. Rodionova for her great help. We would also like to express our gratitude to all the people who have helped and supported us to finish this article.

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. Ziemski M., Junczys-Dowmunt M., Pouliquen B. Parallel Corpus United Nations. *Language Resources and Evaluation (LREC'16)*, Portorozh, Sloveniya, 2016.
2. Hochreiter S., Schmidhuber J. Long Short Term Memory. *Neural Computation*, 1997, Vol. 9, No. 8, Pp. 1735–1780.
3. Gers F.A., Schraudolph N.N., Schmidhuber J. Learning precise timing with LSTM recurrent networks. *Journal of Machine Learning Research*, 2002, Vol. 3, Pp. 115–143.
4. Schuster M., Paliwal K.K. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 1997, Vol. 45, Pp. 2673–2681.
5. Graves A., Mohamed Abdel-Rahman, Hinton G. Speech recognition with deep recurrent neural networks. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, Pp. 6645–6649, preprint arXiv: 1303.5778 03.2013 (<https://arxiv.org/pdf/1303.5778.pdf>).
6. Graves A., Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 2005, Vol. 18, Issue 5–6, Pp. 602–610.
7. Minh-Thang Luong, Hieu Pham, Christopher D. Manning Bilingual word representations with monolingual quality in mind. *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, 2015, Pp. 151–159, preprint arXiv: 1508.04025 08.2015 (<http://www.aclweb.org/anthology/W15-1521>).
8. Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio Neural machine translation by jointly learning to align and translate, preprint arXiv: 1409.0473, 2014.
9. Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, Yoshua Bengio. How to construct deep recurrent neural networks, preprint arXiv: 1312.6026, 12.2013.
10. Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation, preprint arXiv: 1406.1078, 2014.
11. Sutskever I., Vinyals O., Le. Quoc V. Sequence to sequence learning with neural networks. *Proceedings of the Advances in Neural Information Processing Systems*, 2014, Pp. 3104–3112, preprint arXiv: 1409.3215S.
12. Kalchbrenner N., Blunsom Ph. Recurrent continuous translation models. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2013, Pp. 1700–1709.
13. Papineni K., Roukos S., Ward T., Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, 2002, Pp. 311–318.
14. Jean S., Kyunghyun Cho, Memisevic R., Bengio Y. On using very large target vocabulary for neural machine translation. *Conference ACL-2015*, preprint arXiv: 1412.2007, 2014.
15. Minh-Thang Luong, Sutskever I., Quoc V. Le, Vinyals O., Wojciech Zaremba. Addressing the rare word problem in neural machine translation, preprint arXiv: 1410.8206, 2014.
16. Sennrich R., Haddow B., Birch A. Improving neural machine translation models with Monolingual Data. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, preprint arXiv: 1511.06709, 2015.
17. Cheng Y., Xu W., He Z., He W., Wu H., Sun M., Liu Y. Semi-supervised learning for neural machine translation, preprint arXiv: 1606.04596, 2016.
18. Dong D., Wu H., He W., Yu D., Wang H. Multi-task learning for multiple language translation. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 2015, Vol. 1, Pp. 1723–1732.
19. Wang M., Lu Z., Li H., Liu Q. Memory-enhanced decoder for neural machine translation, preprint arXiv: 1606.02003, 2016.
20. Sennrich R., Haddow B. Linguistic input features improve neural machine translation. *Proceedings of the First Conference on Machine Translation*, 2016, Pp.83–91, preprint arXiv: 1606.02892, 2016.
21. Tu Z., Lu Z., Liu Y., Liu X., Li H. Modeling coverage for neural machine translation. *ACL Conference*, preprint arXiv: 1601.04811, 2016.
22. Shen S., Cheng Y., He Z., He W., Wu H., Sun M., Liu Y. Minimum risk training for neural machine translation, preprint arXiv: 1512.02433, 2015.
23. Cohn T., Duy C., Hoang V., Vymolova E., Yao K., Dyer C., Haffari G. Incorporating structural alignment biases into an attentional neural translation model, preprint arXiv: 1601.01085, 2016.
24. Hitschler J., Schamoni S., Riezler S. Multimodal Pivots for Image Caption Translation. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin, Germany, 2016, Pp. 2399–2409.
25. Britz D., Goldie A., Minh-Thang Luong, Quoc Le. Massive exploration of neural machine translation architectures, preprint arXiv:1703.03906, 2017.

Received 08.05.2018. / Статья поступила в редакцию 08.05.2018.

THE AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

TIAN Zhaolin

ТЯНЬ Чжаолинъ

E-mail: peter0431peter@gmail.com

ZHANG Weiwei

ЧЖАН Вэйвэй

E-mail: soszhang@outlook.com

DOI: 10.18721/JCSTCS.11206

UDC 004.81

COMPARISON ANALYSIS OF KNOWLEDGE-BASED SYSTEMS FOR NAVIGATION OF MOBILE ROBOT AND COLLISION AVOIDANCE WITH OBSTACLES IN UNKNOWN ENVIRONMENT

V.N. Sichkar

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

Developing systems for intelligent navigation is one of the major problems in world of modern robotics. This problem is particularly urgent when the environment is unknown. It means that a mobile robot meeting unpredictable obstacles on its way and has to react according to the current situation fast and in real time. That is why developing such a system is always a big challenge. This paper studies different techniques for storing and using the knowledge in order to avoid collisions with obstacles. Most attention is paid for developing two types of Knowledge Bases to help the mobile robot to avoid possible collisions and continue its way. A comparison analysis is provided for these two different types of Knowledge Bases. The advantages and disadvantages were analyzed and described.

Keywords: mobile robot, intelligent navigation, obstacle avoidance, symbolic knowledge base, neural network knowledge base.

Citation: Sichkar V.N. Comparison analysis of knowledge-based systems for navigation of mobile robot and collision avoidance with obstacles in unknown environment. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems, 2018, Vol. 11, No. 2, Pp. 64–73. DOI: 10.18721/JCSTCS.11206

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СИСТЕМ, ОСНОВАННЫХ НА ЗНАНИЯХ ДЛЯ НАВИГАЦИИ МОБИЛЬНОГО РОБОТА И ПРЕДОТВРАЩЕНИЯ СТОЛКНОВЕНИЙ С ПРЕПЯТСТВИЯМИ В НЕИЗВЕСТНОЙ СРЕДЕ

В.Н. Сичкар

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

Разработка систем интеллектуальной навигации — одна из основных проблем в мире современных роботов. Окружающая среда, в которой будет работать мобильный робот, не постоянна. Он будет встречать препятствия на своем пути и при этом должен реагировать в соответствии с текущей ситуацией быстро и в реальном времени. В статье рассмотрены различные методы хранения и использования знаний для того, чтобы робот смог избежать столкновений с препятствиями. Особое внимание уделено разработке двух разных типов баз знаний, на основе которых мобильный робот избегает возможных столкновений с препятствиями и продолжает свой путь. Проведен сравнительный анализ двух разных типов баз знаний. Проанализированы и описаны их преимущества и недостатки.

Ключевые слова: мобильный робот, интеллектуальная навигация, предотвращение препятствий, символьная база знаний, нейросетевая база знаний.

Ссылка при цитировании: Сичкар В.Н. Сравнительный анализ систем, основанных на знаниях для навигации мобильного робота и предотвращения столкновений с препятствиями в неизвестной среде // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2018. Т. 11. № 2. С. 64–73. DOI: 10.18721/JCSTCS.11206

Introduction

Creating an autopilot for cars, autonomous vehicles, mobile robots, rescue robots, etc., is currently an important task. Autopilot systems were first developed exclusively for aircraft, but later automatic control systems were also switched to land vehicles. Nowadays, there are three main types of navigation algorithms for movement control of mobile robots in an unknown environment [1]:

- algorithm based on separation of functions for processing the information received in the process of “recognition-modeling-planning-action”;
- algorithm based on the strategy of targeted behavior of the mobile robot, which includes training under supervision, fuzzy logic, neural networks, and behavior planning based on the data obtained from sensors;
- hybrid system based on integration of the two previous types of algorithms, which

allows to overcome difficulties in modeling and ensure adaptability of behavior in real time and in uncertain environment.

Navigation in the field of mobile robotics also can be classified in two types: global navigation and local navigation [2]. In the part of global navigation, the preliminary knowledge of the environment should be available. With the help of the local navigation the mobile robot can orient itself using ultrasonic sensors, camera, lidar sensors, and variety of other sensors according to the real task.

Problem statement

The study is aimed at comparing different types of Knowledge Bases for navigation system of a mobile robot in an unknown environment in order to avoid possible collisions with obstacles. For comparison analysis, two types of Knowledge Bases were chosen and they are a Symbolic Knowledge Base and a Neural Net-

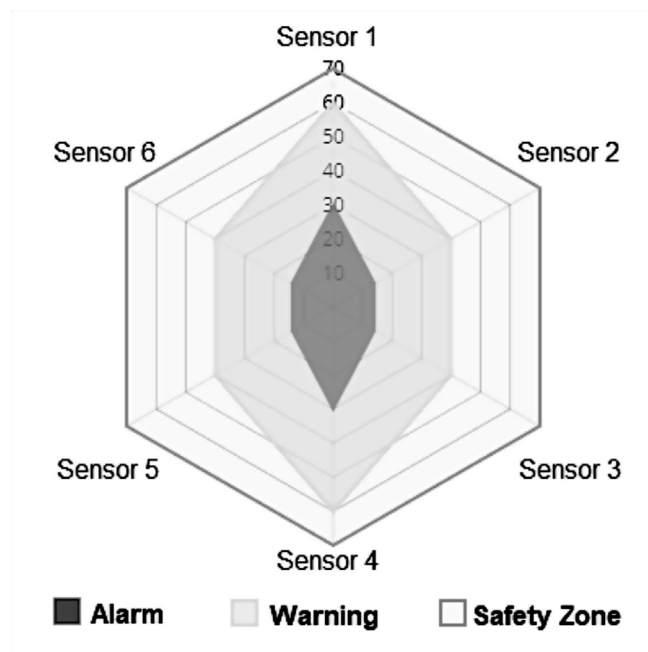


Fig. 1. Radar chart with zones

work Knowledge Base. This study shows the implementation of two types of Knowledge Bases and provides analysis of advantages and disadvantages of both of them.

A mobile robot acts in unknown environment inside a building with static obstacles such as furniture, walls and doors. The surface of the floor is smooth with a small number of low obstacles like thresholds in door frame which are not higher than two centimeters.

The navigation system uses six ultrasonic sensors that monitor different directions while the mobile robot is moving. The type of the sensors is HC-SR04 that can measure distance in the range from 2 to 400 cm and with a step of 1 cm. This range is sufficient for recognition of obstacles in the room. Based on this data (types of obstacles and features of sensors), three zones were developed: Alarm, Warning and Safety. Each zone covers its own space around the mobile robot. Fig. 1 shows a radar chart with zones and Table 1 shows intervals for each of three zones.

The sensors are located along the perimeter of the mobile robot creating a kind of bubble around it. System sends requests to the Knowledge Bases and asks for the current statement for all sensors each time when an obstacle penetrates inside this bubble. For getting results, system uses two types of the Knowledge Bases in parallel. Each sensor works independently and shows in real time the current value which is the distance to the nearest object. The navigation system processes data from all sensors and with the help of Knowledge Base obtains the three statements in each of the six directions. These statements are safety zone, warning zone and alarm zone. After getting this information,

the system slows down the mobile robot and applies an algorithm for maneuvering to avoid the obstacle, or completely stop the mobile robot if there is not enough space to turn left or right.

As shown in Table 1, the range for the alarm zone is between 2 and 30 cm for sensors number 1 and 4 that are front and back sensors, and between 2 and 14 cm for four side sensors. The minimum distance was set to 2 cm due to the blind zone of the chosen ultrasonic sensors. It means that if an object approaches closer, a collision happens. Since the step with which the sensors can measure distance is 1 cm, the three zones have a gap of 1 cm between each other. That is why, for example, the alarm zone for sensor 1 ends in the 30th cm and warning zone begins with the 31st cm. There is no maximum limit for the safety zone because the system does not consider everything that is far from the warning zone as an obstacle.

Symbolic Knowledge Base

The Symbolic Knowledge Base can be shown graphically and parametrically. For the parametric method of representation, the Semantic Web Language (SWL) is used. This technique allows not just to save the database but to show the relationships between the data. In symbolic representation by SWL, the Knowledge Base consists of ontology [3, 4] which includes individuals and their properties. Ontology describes state-independent information, the logical component concept model with particular syntax, ontology class graph. The core of this Knowledge Base contains state-independent information, the actual data component contains all individual instances,

Table 1

Intervals for each zone

Sensor	Alarm min, cm	Alarm max, cm	Warning min, cm	Warning max, cm	Safety, cm
Sensor 1	2	30	31	60	> 61
Sensor 2	2	14	15	40	> 41
Sensor 3	2	14	15	40	> 41
Sensor 4	2	30	31	60	> 61
Sensor 5	2	14	15	40	> 41
Sensor 6	2	14	15	40	> 41

the ontology instance graph. To use information from the ontology instance graph, the Resource Description Framework (RDF) triples [5] are used. For this study, the RDF store was developed with triples in order to have an opportunity to query the Knowledge Base and get the resulting information from it. The queries will be provided by SPARQL query language [6] and the system will get the current statement of each ultrasonic sensor in real time.

A flowchart with a graphical representation of the Knowledge Base is shown below in Fig. 2. For this flowchart, the graphical tool was used and it is called Concept Map. It helps to organize and represent knowledge by concepts with boxes or circles of specific types. Relationships between concepts are connected by line and words on the line. The words or phrases specify the relationship between the two concepts [7].

This flowchart describes the intervals in which the system has Alarm and Warning statements. Alarm and Warning here are the concepts that are connected with the State concept by the linking phrase «a kind of» (AKO). These two concepts describe statements of the system and they are connected with the

concepts of intervals by the appropriate sensor. These sensors are represented here as P1, P2, P3, P4, P5 and P6. Each interval has its own value for maximum and minimum. The intervals are different for Alarm and Warning concepts. Using the Symbolic Knowledge Base is easy because it is easy to change values, add more concepts and link them all together.

Neural Network Knowledge Base

There is another way how to store and query the Knowledge Base of the system by using the Knowledge-Based Neural Network. It means that all data of the system will be inside the Neural Network and asking for the current state of the system is done through the Neural Network. The Neural Network for the system is shown in Fig. 3 below.

The designed Neural Network has an input layer, a first layer, one hidden layer and an output layer. The input layer represents signals from sensors S1–S6 and sends this information to neurons N1–N24 in the first layer. The first layer processes the received data and sends the results to neurons H1–H12 in the hidden layer. The hidden layer also processes data and sends it further to the final output layer. The chosen

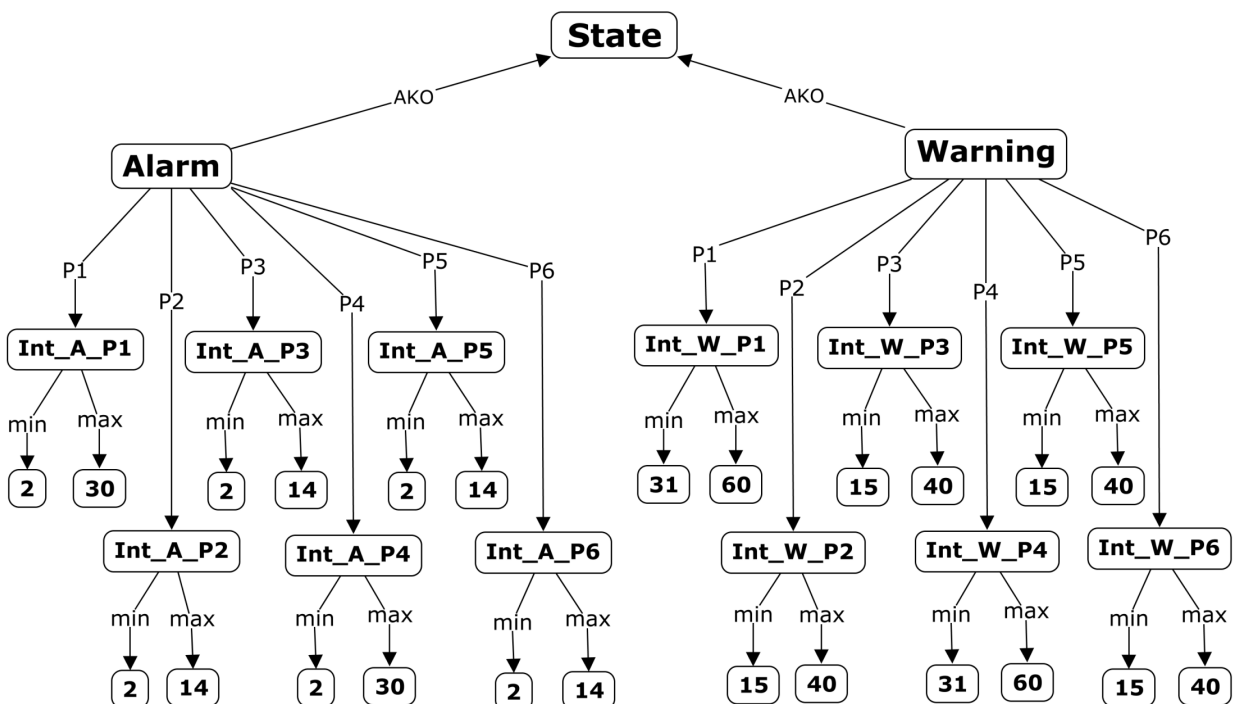


Fig. 2. Flowchart of Knowledge Base by Concept Map

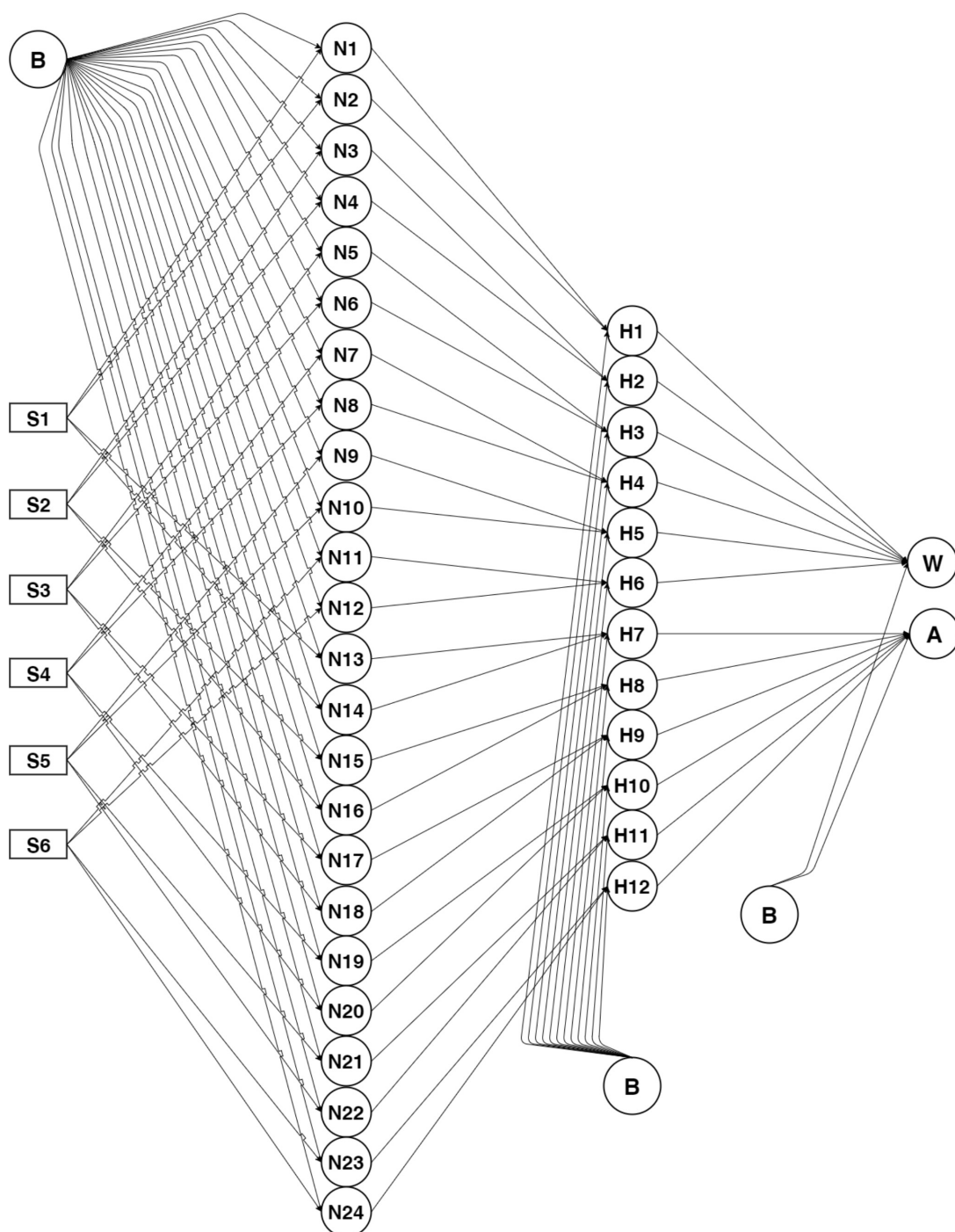


Fig. 3. Neural Network Knowledge Base

number of neurons in the first layer and the added hidden layer are needed because of the complexity of the conditions for calculating and finding out one from two statements for each sensor, Alarm or Warning. The output layer activates one of two neurons if at least

one sensor was in the critical interval. Each layer has an extra neuron called «bias» and it is marked as letter B on the scheme. The bias is an artificial threshold with its specific value for each connection and it plays the role of a kind of «helper» to adjust weights giving more

or less significance to the appropriate neuron. Bias is always activated. To provide comparison analysis for querying time with the Symbolic Knowledge Base, the output layer was improved to show the results for each sensor separately.

The main specific characteristic of the Knowledge-Based Neural Network is that the knowledge is put inside in advance by adjusting the appropriate weights and there is no need to train the Neural Network and it can be used straight away. This can be done if the conditions are known from the very beginning.

The character of relationships in the developed Neural Network is a feed forward. It means that all communication is directed strictly from input neurons to output ones [8–13].

Algorithm of the system operation

The primary task in constructing the algorithms for overcoming obstacles for the mobile robot is intellectualization and automation of the control processes themselves. The mobile robot has to perform the task of navigation in an unknown environment. The goal is to explore the unknown environment, collect data about the distances from the ultrasonic sensors, send them to the central computer and come back to the initial point.

The mobile robot is equipped with ultrasonic sensors that will be used for the decision-making process. At the same time, the operator will be able to control the mobile robot at a distance by remote control if some urgent situation happens [14, 16]. The sensors will read

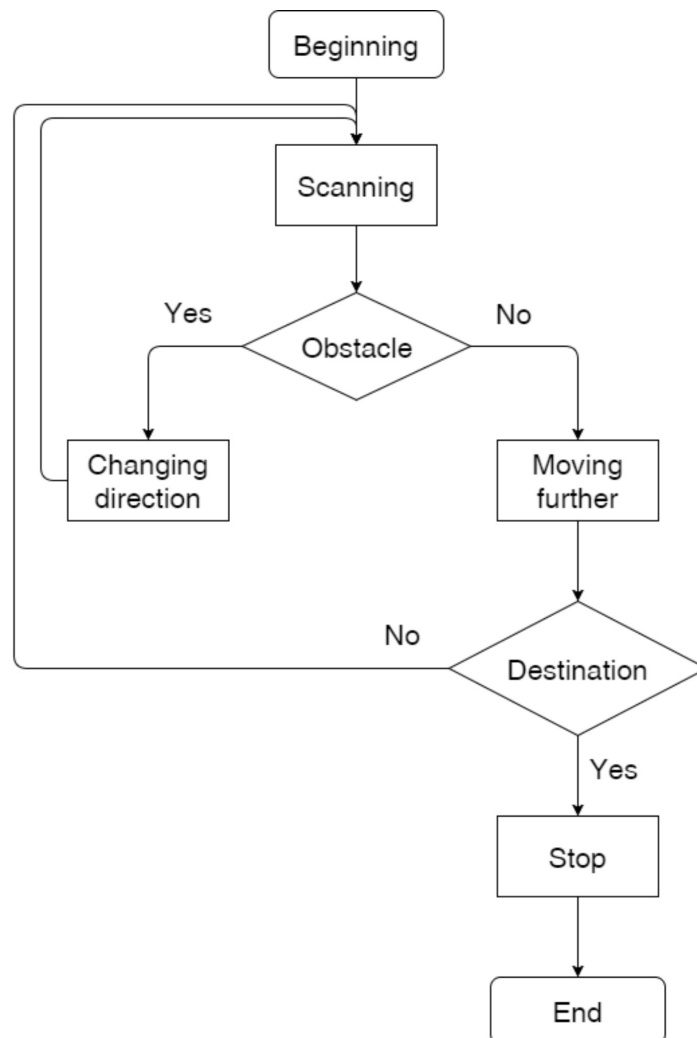


Fig. 4. Flowchart of the algorithm for overcoming an obstacle

the data from the environment and record the distance to the objects around the mobile robot. If there is a probability of a collision or the current situation may cause an accident, the system slows down the mobile robot and runs the obstacle-overcoming algorithm.

Fig. 4 shows a simplified algorithm for the movement of the mobile robot.

Equipment

A six-wheel base, namely, 6WD Smart Carwas chosen as the mobile robot for the implementation of the hardware part of the project. It has active suspension that helps to go in rough terrain. Also, the advantage of this mobile robot is in its ability to carry heavy equipment. The own weight of the mobile robot is about 2 kg and it can carry about 3 kg, which is enough for all needed sensors, microcontrollers, batteries, wires, etc. Other parameters:

- size: 28x21x11.5 cm;
- weight: 2 kg;
- maximum load: 3 kg;
- number of DC motors: 6;
- working voltage: 12 v.

Ultrasonic sensors HC-SR04 were chosen

as sensors for scanning the environment around the mobile robot. Selection of this type of ultrasonic sensors is due to their low cost and quite good quality for the purposes of scientific experiments. They have a good accuracy, which is 1 cm, and the distance of measurement up to 4 m. All these features and six sensors together make it possible to carry out scanning of the environment around in 360 degrees and recognize necessary objects and detect obstacles. The main disadvantage of this type of ultrasonic sensors is that if the environment is too dusty, it can cover the sensors and make it difficult for them to work properly.

To collect data received from the sensors, a single-board microcontroller Arduino Mega is used. All data are sent to the central computer which processes information. Exchange of information is provided via a Bluetooth module. Also, Arduino Mega controls the six motors by sending commands to move and directions to go.

Implementation results

For the implementation of the theoretical part of the system, the C Sharp Windows Form application was developed. As a result,

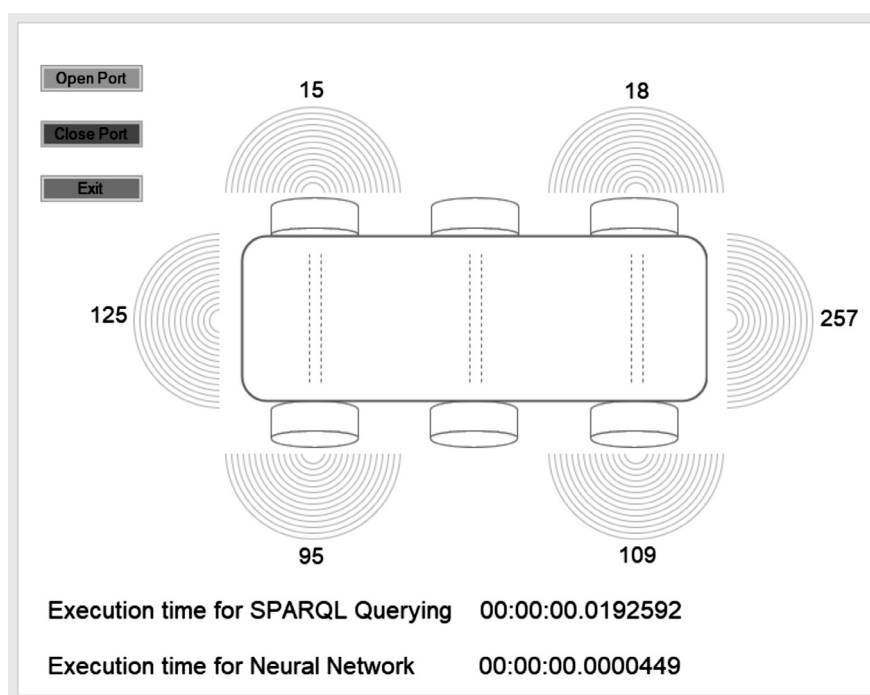


Fig. 5. C Sharp Windows Form application for representing the results



the application shows the system working in real time with an interface that is intuitively understandable for the operator. The application is directly connected with the mobile robot via a Bluetooth connection providing transmission of data in both ways. Fig. 5 shows the main window of the application.

As shown in the main window, the operator can see and control the zones around the mobile robot in six different directions. If the area is clear and there is no Warning or Alarm statements, the system allows the mobile robot to enter and pass through the checked area. The values from all ultrasonic sensors are being checked continuously 5 times a second. This feature provides enough data for the algorithm of the system and based on the results it makes a decision which allows to avoid any collision.

All data received from the sensors are being processed in two different threads in parallel. One thread is developed for querying the Symbolic Knowledge Base and another one is developed for querying the Neural Network Knowledge Base. This approach allows to check and compare the execution time for each thread. In the program, the Symbolic Knowledge Base is loaded as a graph and querying is done by SPARQL language through this graph. The Neural Network Knowledge Base is represented by matrices in the program and results are being obtained by sending data to the input layer.

As seen in the main window in Fig. 5, the processing time is shown for both threads in

real time. Comparison analysis was provided by collecting information about execution times from two threads for two different types of Knowledge Bases. The results were evaluated through a variety of tests in different conditions for the mobile robot. Table 2 shows the execution time in ten experiments.

These experiments were done in different situations. Experiment number 1 was done with no obstacle detected in the Warning zone, and the result shows that the execution time is minimum for both types of Knowledge Bases. Experiment number 2 was done when all sensors detected an obstacle in the Warning zone. Although the situation when the mobile robot is completely surrounded with obstacles was not possible for current research, it was important to understand how fast the system will react on this condition. Other experiments were done by different combinations with the number of obstacles in different zones, and these conditions are described below.

Experiment 3: side sensors number 3 and 6 detected obstacles in Warning zone, back sensor number 4 detected obstacle in Alarm zone; other sensors did not detect obstacles in Alarm and Warning zones.

Experiment 4: side sensors number 2 and 6 detected obstacles in Alarm zone, front sensor number 1 detected an obstacle in Warning zone; other sensors did not detect obstacles in Alarm and Warning zones.

Experiment 5: side sensors number 5 and 6 detected obstacles in Warning zone in

Table 2

Testing results for the execution time for two different Knowledge Bases

Experiment	Symbolic Knowledge Base	Neural Network Knowledge Base
1	0.0125929	0.0000458
2	0.9172391	0.0009884
3	0.1740155	0.0005356
4	0.2929742	0.0000854
5	0.5692983	0.0000973
6	0.0926548	0.0003466
7	0.8287374	0.0009266
8	0.0284718	0.0000774
9	0.0413314	0.0005626
10	0.0155394	0.0003917

the border close to Alarm zone, back sensor number 4 detected an obstacle in Alarm zone; other sensors did not detect obstacles in Alarm and Warning zones.

Experiment 6: front and back sensors number 1 and 4 detected obstacles in Alarm zone; other sensors did not detect obstacles in Alarm and Warning zones.

Experiment 7: side sensors number 2, 3, 5 and 6 detected obstacles in the Alarm zone; other sensors did not detect obstacles in Alarm and Warning zones.

Experiment 8: side sensors number 3 and 5 detected obstacles in Warning zone; other sensors did not detect obstacles in Alarm and Warning zones.

Experiment 9: side sensors number 2 and 3, front sensor number 1 detected obstacles in the Warning zone; other sensors did not detect obstacles in Alarm and Warning zones.

Experiment 10: only one side sensor number 6 detected an obstacle in Alarm zone; other sensors did not detect obstacles in Alarm and Warning zones.

Experiments number 1 and 2 showed a strong dependence between detecting the obstacles and the execution time for both two types of Knowledge Bases. When no obstacle was detected in experiment number 1, the execution time was one order less for both threads than it was in experiment number 2, when all sensors detected obstacles in Warning zone. Experiment number 2 showed the maximum execution time over all 10 experiments.

Experiments from 3 to 10 showed that a different number of obstacles detected by the sensors has a different effect on the execution time for both threads, and can vary due to the zone in which the obstacle was detected. These experiments revealed that it takes more time to process data for an obstacle detected in Warning zone in the border with Alarm zone than for an obstacle detected in Alarm zone. This situation occurs because requests are sent first to check if the obstacle was detected in Warning zone close to Alarm zone, and if it was, then requests are sent further to check if the obstacle was detected in Alarm zone also. That is why execution time is longer for cases when the obstacle is in the border between Warning and Alarm zones.

Execution time is longer when more obstacles are detected because each sensor that detects obstacles in Warning or Alarm zones starts to send data to be processed. A minimum difference of around 500 times between the execution time for two different approaches was recorded. A maximum difference of around 11,000 times was recorded. Absolutely in all cases the thread with the Neural Network Knowledge Base was much faster, which is a big advantage for using this approach in real time systems with the ability for fast decision-making.

Conclusions

In this paper, the system that controls the statements around the mobile robot was developed. The system has six sensors that control the environment around the mobile robot creating a “bubble” and if there is any penetration inside this “bubble”, then system reacts immediately. The system automatically takes control of the mobile robot if there is a probability of collision.

The main research was put into two different types of Knowledge Bases that are implemented in the system. Each of them have advantages and disadvantages. The main advantage of the Symbolic Knowledge Base is in the ease of creating, adding and changing data inside it. However, the comparison analysis made on a series of experiments showed that the Neural Network Knowledge Base was always much faster and the difference was in the range from 500 to 11,000 times. The main advantage of the Neural Network Knowledge Base is in the speed of processing data. The main disadvantage of this type of Knowledge Base is in the necessity to adjust the weights inside the Neural Network each time it is needed to be correct and it also makes it more complicated to add and update data.

For the systems where data is stable and there is no need to update it too often, the Neural Network Knowledge Base is more suitable. For the systems where data has to be updated very often, the Symbolic Knowledge Base is more suitable as it can be done easier. Also, the processing time has to be considered, as the Neural Network Knowledge Base gives the results much faster.



REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. Rao A., Elara M.R., Elangovan K. Constrained VPH+: a local path planning algorithm for a bio-inspired crawling robot with customized ultrasonic scanning sensor. *Robotics and Biomimetics*, 2016, Vol. 3, Pp. 1–13. DOI: 10.1186/s40638-016-0043-1
2. Yakoubi M., Laskri M. The path planning of cleaner robot for coverage region using Genetic Algorithms. *Journal of Innovation in Digital Ecosystems*, 2016, Vol. 3, Issue 1, Pp. 37–43. DOI: 10.1016/j.jides.2016.05.004
3. Feilmayr C., Wolfram W. An analysis of ontologies and their success factors for application to business. *Data & Knowledge Engineering*, 2016, Vol. 101, Pp. 1–23. DOI: 10.1016/j.datak.2015.11.003
4. Labra Gayo J.E., Jeuring J., Alvarez Rodriguez J.M. Inductive triple graphs: A purely functional approach to represent RDF. *Graph Structures for Knowledge Representation and Reasoning*, 2014, Vol. 8323, Pp. 92–110. DOI: 10.1007/978-3-319-04534-4_7
5. Yuan P., Xie C., Jin H. Dynamic and fast processing of queries on large-scale RDF data. *Knowledge and Information Systems*, 2014, Vol. 41, Issue 2, Pp. 311–334. DOI: 10.1007/s10115-013-0726-7
6. McCarthy L., Vandervalk B., Wilkinson M. SPARQL Assist language-neutral query composer. *BMC Bioinformatics*, 2012, Vol. 13, Pp. 1–9. DOI: 10.1186/1471-2105-13-S1-S2
7. The theory underlying concept maps and how to construct and use them. Available: <https://cmap.ihmc.us/docs/theory-of-concept-maps> (Accessed: 23.01.2017).
8. Sailamul P., Jang J., Paik S.B. Synaptic convergence regulates synchronization-dependent spike transfer in feedforward neural networks. *Journal of Computational Neuroscience*, 2017, Vol. 43, Issue 3, Pp 189–202. DOI: 10.1007/s10827-017-0657-5
9. Arulmozhi V., Reghunadhan R. Predicting the protein localization sites using artificial neural networks. *Journal of Cheminformatics*, 2013, Vol. 5, Issue 1, P. 46. DOI: 10.1186/1758-2946-5-S1-P46
10. Zhang W., Cao J., Wu R. Projective synchronization of fractional-order delayed neural networks based on the comparison principle. *Advances in Difference Equations*, 2018, Vol. 73, Pp. 1–16. DOI: 10.1186/s13662-018-1530-1
11. Gunther F., Pigeot I., Bammann K. Artificial neural networks modeling gene-environment interaction. *BMC Genetics*, 2012, Vol. 13, Pp. 1–18. DOI: 10.1186/1471-2156-13-37
12. Zhang Y., Yamaguchi R., Imoto S. Sequence-specific bias correction for RNA-seq data using recurrent neural networks. *BMC Genomics*, 2017, Vol. 17, Issue 11, Pp. 1–6. DOI: 10.1186/s12864-016-3262-5
13. Al Machot F., Ali M., Haj Mosa A. Real-time raindrop detection based on cellular neural networks for ADAS. *Journal of Real-Time Image Processing*, 2016, Vol. 1, Pp. 1–13. DOI: 10.1007/s11554-016-0569-z
14. Kobayashi Y., Hosoe S. Cooperative enclosing and grasping of an object by decentralized mobile robots using local observation. *International Journal of Social Robotics*, 2012, Vol. 4, Issue 1, Pp. 19–32. DOI: 10.1007/s12369-011-0118-7
15. Dennis L.A., Fisher M., Lincoln N.K. Practical verification of decision-making in agent-based autonomous systems. *Automated Software Engineering*, 2016, Vol. 23, Issue 3, Pp. 305–359. DOI: 10.1007/s10515-014-0168-9
16. Lunenburg J., Molengraft R., Steinbuch M. A representation method based on the probability of collision for safe robot navigation in domestic environments. *Autonomous Robots*, 2018, Vol. 42, Issue 3, Pp 601–614. DOI: 10.1007/s10514-017-9653-x

Received 19.03.2018. / Статья поступила в редакцию 19.03.2018.

THE AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

SICHKAR Valentin N.
СИЧКАР Валентин Николаевич
E-mail: valenty.n.s2014@yandex.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»

Том 11, № 2, 2018

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*

Перевод на английский язык *А.С. Колгатина*

Лицензия ЛР № 020593 от 07.08.97

Подписано в печать 29.06.2018. Формат 60×84 1/8. Бум. тип. № 1.
Печать офсетная. Усл. печ. л. 8,83. Уч.-изд. л. 8,83. 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 — 4 номера 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