РОССИЙСКИЙ ФОНД ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ
РФФИ

ISSN 1605-8070
eISSN 2410-4639

ВЕСТНИК РФФИ

№4 (92) октябрь–декабрь 2016 г.

ТЕМАТИЧЕСКИЙ БЛОК:
ОБРАБОТКА ИЗОБРАЖЕНИЙ
И РАСПОЗНАВАНИЕ ОБРАЗОВ

стр.
17

Вестник Российского фонда фундаментальных исследований

№ 4 (92) октябрь–декабрь 2016 года

Основан в 1994 году

Зарегистрирован Комитетом РФ по печати, рег. № 012620 от 03.06.1994

Сетевая версия зарегистрирована Роскомнадзором, рег. № ФС77-61404 от 10.04.2015

Учредитель

Федеральное государственное бюджетное учреждение «Российский фонд фундаментальных исследований»

Адрес редакции:

119334, г. Москва, Ленинский проспект, 32а

Тел.: (499) 995-16-05

e-mail: pressa@rfbr.ru

Russian Foundation for Basic Research Journal

N 4 (92) October-December 2016

Founded in 1994

Registered by the Committee of the Russian Federation for Printed Media, 012620 of 03.06.1994 (print)

Registered by the Roskomnadzor FS77-61404 of 10.04.2015 (online)

The Founder

Federal State Institution

"Russian Foundation for Basic Research"

Editorial address:

32a, Leninskiy Ave., Moscow, 119334, Russia

Tel.: (499) 995-16-05

e-mail: pressa@rfbr.ru

«Вестник РФФИ»
№ 4 (92) октябрь–декабрь 2016 г. (Приложение к «Информационному бюллетеню РФФИ» № 24)

КОЛОНКА ТЕМАТИЧЕСКОГО РЕДАКТОРА

ТЕМАТИЧЕСКИЙ БЛОК: ОБРАБОТКА ИЗОБРАЖЕНИЙ И РАСПОЗНАВАНИЕ ОБРАЗОВ

"RFBR Journal"
N 4 (92) October-December 2016 (Supplement to "Information Bulletin of RFBR" N 24)

THEMED ISSUE EDITOR'S COLUMN

THEMED SECTION: IMAGE PROCESSING AND PATTERN RECOGNITION

О редакторе тематического блока академике РАН, профессоре С.В. Емельянове

- *Руководитель секции информационных технологий и автоматизации отделения нанотехнологий и информационных технологий (ОНИТ) РАН*
- *Заместитель академика секретаря ОНИТ РАН*
- *Научный руководитель Федерального исследовательского центра «Информатика и управление» РАН (ФИЦ ИУ РАН)*
- *Член Института инженеров электротехники и электроники (Institute of Electrical and Electronics Engineers, IEEE)*
- *Член Нью-Йорской академии наук*

- *Head of the Section of Information Technologies and Automatization of the Department of Nanotechnologies and Information Technologies (DNIT), RAS*
- *Deputy Academician-Secretary of DNIT, RAS*
- *Scientific Director of the Federal Research Center "Computer Science and Control" RAS (FRC CSC RAS)*
- *Member of IEEE, Member of the New York Academy of Sciences*
- *Foreign Member of the Academy of Sciences of Bosnia and Herzegovina*
- *Honorary Doctor of the Sarajevo University*

- *Иностранный член Академии наук Боснии и Герцоговины*
- *Почетный доктор университета г. Сараево*
- *Заведующий кафедрой нелинейных динамических систем и процессов управления факультета вычислительной математики и кибернетики Московского государственного университета им. М.В. Ломоносова (МГУ)*
- *Член редакционных коллегий, редакционных советов ряда журналов:*

  «Автоматика и телемеханика», «Дифференциальные уравнения», «Доклады РАН», «Информационные технологии и вычислительные системы», «Искусственный интеллект и принятие решений», «Информатика и ее применение»
- *Председатель Совета по математике при Министерстве образования РФ*
- *Член ученых и специализированных советов при МГУ и ФИЦ ИУ РАН*

- *Director of the Department of Nonlinear Dynamic Systems and Control Processes of Computational Mathematics and Cybernetics Faculty of Lomonosov Moscow State University (MSU)*
- *Member of editorial boards of several journals:*

  "Automation and Remote Control", "Differential equations", "Doklady of the Academy of Sciences", "Information technologies and computational systems", "Artificial intelligence and decision-making", "Informatics and its application"
- *Chairman of the Council for Mathematics of the Ministry of Education of the Russian Federation*
- *Member of Science and Academic Councils of MSU and FRC CSC RAS.*

Государственные награды, звания и премии:

- *Ленинская премия (1972)*
- *Орден Октябрьской Революции (1974)*
- *Орден Дружбы народов (1979)*
- *Государственная премия СССР (1980)*
- *Орден «За заслуги» (Польша, 1989)*
- *Орден Кирилла и Мефодия I степени (Болгария, 1990)*
- *Российской Федерации в области науки и техники (1994)*
- *Премия Президиума РАН им. академика А.А. Андронова (2000)*
- *Лауреат Ломоносовской премии МГУ по науке I степени (2002)*
- *Орден Почета (2010)*
- *Орден «За заслуги перед Отечеством» II (2015), III (2004), IV (1999) степени*

Honours and awards:

- *Lenin Prize (1972)*
- *Order of the October Revolution (1974)*
- *Order of the Friendship of Peoples (1979)*
- *USSR State Prize (1980)*
- *Cross of Merit (Poland, 1989)*
- *Order of Cyril and Methodius, 1st class (Bulgaria, 1990)*
- *State Prize of the Russian Federation in Science and Technology (1994)*
- *Academician Andronov Prize of the Presidium of RAS (2000)*
- *Lomonosov MSU Science Prize, 1st class (2002)*
- *Order of Honour of the Russian Federation (2010)*
- *Order "For Merit for the Motherland", 2nd class (2015), 3rd class (2004), 4th class (1999)*

Емельянов Станислав Васильевич учился в Московском авиационном институте на факультете приборостроения и систем управления летательных аппаратов (1947–1952), затем (1953–1957) в аспирантуре (без отрыва от производства) при Институте автоматики и телемеханики АН СССР (ныне Институт проблем управления РАН).

В 1952 г. С.В. Емельянов поступил на работу в Институт автоматики и телемеханики, где прошел путь от инженера до заместителя директора по науке (заместитель директора с 1967 по 1975 г.). С 1976 г. работает в Институте системных исследований АН СССР (в настоящее время Институт системного анализа РАН – ИСА РАН). С 1993 по 2003 г. – директор ИСА РАН; с 1977 по 2002 г. – генеральный директор Международного научно-исследовательского института проблем управления. В настоящее время С.В. Емельянов – научный руководитель Федерального исследовательского центра «Информатика и управление».

Заведующий кафедрой нелинейных динамических систем и процессов управления факультета вычислительной математики и кибернетики (с 1989 г.). Почетный профессор Московского государственного университета им. М.В. Ломоносова (МГУ) (1998), заслуженный профессор МГУ (1999).

Кандидат технических наук (1958), тема диссертации – «Системы автоматического управления с переменной структурой». Доктор технических наук (1964), тема диссертации – «Теория систем с переменной структурой»; профессор (1966).

Член-корреспондент АН СССР (1970), действительный член РАН (1984), академик-секретарь Отделения информатики, вычислительной техники и автоматизации РАН (1992–2002).

Является автором 18 книг и свыше 250 статей, 70 патентов на изобретения в области теории и практики управления динамическими системами, их анализа и оптимизации.

С.В. Емельянов – основатель известной научной школы. Он подготовил более 30 докторов и 70 кандидатов наук; среди его учеников – академики и члены-корреспонденты РАН, члены других академий, руководители институтов, фирм.

About the Editor of the Themed Section Academician, Professor S.V. Emelyanov

Stanislav Vasilevich Emelyanov was studying in Moscow Aviation Institute, Instrumentation and Control Systems of Aircraft Faculty in 1947–1952. In 1953–1957 he was a graduate student at the Institute for Automatics and Telemechincs AS USSR (V.A. Trapeznikov Institute of Control Sciences of RAS nowadays).

In 1952 S.V. Emelyanov had started his career in the Institute for Automatics and Telemechanics as an engineer and later had grown up to Deputy Director for Science (1967-1975). In 1976 S.V. Emelyanov was admitted to the Institute for Systems Research AS USSR (the Institute for Systems Analysis RAS – ISA RAS nowadays). From 1977 until 2002 he served as the General Director at the International Research Institute for Advanced Systems (IRIAS), from 1993 until 2003 – as the ISA RAS Director. Currently, S.V. Emelyanov is the Scientific Director at the Federal Research Center "Computer Science and Control" (FRC CSC RAS).

Academician S.V. Emelyanov also holds the positions of the Director of the Department of Nonlinear Dynamic Systems and Control Processes of Computational Mathematics and Cybernetics Faculty (since 1989), Honorary Professor (since 1998) and Emeritus Professor (since 1999) of Lomonosov MSU.

He defended his PhD thesis for the Degree of Candidate of Technical Sciences in 1958 (the topic was "Automated control systems with variable structure"), and in 1964 he had got his Doctor of Technical Science Degree (his dissertation theme was "Theory of systems with variable structure"). The academic rank of Professor was conferred on him in 1966.

S.V. Emelyanov had become a Corresponding Member of the Academy of Sciences of USSR in 1970, and full member of RAS – in 1984, he also was the Academician-Secretary for the Department of Informatics, Computational Technology and Automatization of RAS in 1992–2002.

S.V. Emelyanov authored 18 books, more than 250 articles, 70 invention patents on the theory and practice of control, analysis and optimization of dynamic systems.

He was the founder of the well-known scientific school. Under scientific supervision of professor S.V. Emelyanov more than 30 theses for the degree of a Doctor of Science and 70 theses for the degree of a Candidate of Science have been defended; his former students at present are academicians and corresponding members of RAS, members of other academies, heads of institutions and commercial firms.

Видеопоток. Новый взгляд на распознавание ригидных объектов

С.В. Емельянов

Последние 10–15 лет мы наблюдаем беспрецедентный взлет интереса к фундаментальным и прикладным исследованиям, посвященным обработке и распознаванию изображений. Он связан, прежде всего, с бурным развитием цифровой фотографии, позволяющей получать видеоряды, состоящие из фотографий высокого разрешения, на сравнительно недорогих устройствах.

Это заставило совершенно по-новому подойти к решению всех важных задач распознавания, начиная от распознавания разного рода документов и заканчивая распознаванием лиц, сцен или ледовой обстановки.

Во всем мире в эту тематику вкладываются немалые деньги. Это касается как организаций, подобных Национальному научному фонду (National Science Foundation) и Японскому институту физико-химических исследований (RIKEN), поддерживающих чисто научные исследования, так и таких, как Google и Microsoft, финансирующих крупные прикладные проекты гражданского (например беспилотный автомобиль) или военного (например системы оптического наведения) назначения.

Российский фонд фундаментальных исследований в 2013 г. открыл трехлетнюю тему ориентированных междисциплинарных исследований (офи_м) «Восприятие и анализ цветных изображений в видеопотоке и распознавание сложных ригидных объектов». В проектах, проводимых в рамках названной темы, получен ряд фундаментальных результатов, имеющий теоретическое и прикладное значения. По большинству из них в тематическом блоке настоящего выпуска представлены авторские статьи.

По замыслу тема была ограничена распознаванием так называемых ригидных объектов, т.е. объектов, не меняющих своей формы. Например, автомобиль – ригидный объект, а лошадь – нет. Впрочем, как часто бывает, есть много пограничных случаев. Не всегда ясно, куда отнести «изогнутую» страницу раскрытой книги или дерево при сильном ветре. Распознавание таких объектов в видеопотоке имеет важные преимущества по сравнению с распознаванием их на фотографиях. Разумеется, всегда можно распознать первый попавшийся кадр или попытаться выбрать из имеющихся изображений в каком-то смысле наилучшее. Однако это только первый шаг.

При наличии нескольких кадров, содержащих один объект, часто удается построить на всех изображениях единую систему координат. Тогда весь ряд изображений можно рассматривать как набор независимых экспериментов, а множество значений в каждой точке – как реализацию случайной величины. Благодаря этому методы теории вероятностей становятся применимыми к исследованию алгоритмов распознавания без всяких оговорок, обычно сопровождающих такие рассмотрения.

Кроме того, большинство современных видеокамер имеют механизмы управления своими параметрами (фокусом, экспозицией, уровнем подсветки, балансом белого и др.) Это открывает возможности включения обратных связей и позволяет решать задачу распознавания с использованием аппарата теории управления.

В то же время надо понимать, что обработка видеопотока сопровождается рядом технических проблем. Первая группа проблем связана с неконтролируемыми условиями съемки (освещение, ракурс, движение и пр.). Соответственно, используемые модели должны быть инвариантны к возникающим искажениям. Источником проблем второй группы является то, что видеопоток представляет собой огромный массив информации, которая должна обрабатываться в реальном времени. Это требует

создания вычислительно эффективных алгоритмов как собственно распознавания, так и обработки данных. Третья группа проблем связана с обработкой видеокадров, полученных с мобильных устройств, прежде всего, с существенно более низким разрешением, чем на обычных фотоаппаратах или сканерах. Кадр видео имеет, как правило, число точек, равное 2–4 мегапикселям, тогда как камеры фотоаппаратов – десятки мегапикселей.

Кроме того, большую часть практически значимых задач распознавания видеопотока необходимо решать непосредственно на мобильных устройствах. Передача всей информации на мощные сервера обычно невозможна по различным причинам. Вот несколько примеров. Автопилот автомобиля не может тратить время на передачу данных. Спутник не имеет возможности передачи столь большого объема информации из-за отсутствия достаточно широкого канала связи. Автономные подводные аппараты вообще могут передавать только крайне ограниченный объем данных.

Даже в такой сравнительно очевидной задаче, как распознавание паспорта с помощью мобильного устройства, есть серьезные причины распознавания непосредственно на устройстве. Здесь дело в правовых аспектах, а именно в том, что персональные данные должны быть защищены. При этом видеопоток гораздо труднее надежно зашифровать в сравнении с результатами распознавания, составляющими не десятки миллионов, а всего лишь считанные сотни байт.

В рамках данного сборника вопросы распознавания видеопотока рассматриваются с нескольких сторон: использование алгоритмов «глубокого обучения», создание и оптимизация статистических методов распознавания текстов, устойчивых к искажениям, возникающим в условиях неконтролируемой съемки; различные подходы к использованию синтаксических методов распознавания, базирующихся как на статистических свойствах языка, так и на статистической модели распознавания, в видеопотоке; вопросы построения систем распознавания полностью базирующихся на видеопотоке; создание модели искажений в томографии; а также аналитические методы распознавания, опирающиеся на проективные инварианты.

В работе Т.С. Чернова, Д.А. Ильина и др. исследуются два современных метода сегментации печатного текста, устойчивые к различным искажениям, возникающим при съемке с мобильных устройств. Первый предложенный метод развивает классические подходы к сегментации для случая распознавания видеопотока. А наиболее любопытный результат работы – второй метод, широко использующий подходы машинного обучения, в частности сверточные и рекуррентные нейронные сети, что делает возможным создание алгоритмов сегментации без разработки множественных эвристик, привязанных к конкретным особенностям текста.

В статье Е.Г. Кузнецовой, И.В. Полякова и др. исследуется подход к самообучению, основанный на систематическом выявлении ошибок по двум признакам: внутренняя противоречивость серии результатов распознавания и повторяемость таких серий на похожих образах. Интересно, что случаи, аналогичные найденным, предлагается исправлять «на лету» за счет использования поисковой структуры похожих случаев.

Работа Е.Е. Лимоновой, А.В. Шешкуса и др. посвящена вычислительной оптимизации глубоких нейронных сетей за счет аппроксимации сверхточных фильтров и использованию целочисленной арифметики. Благодаря применению линейной комбинации сепарабельных фильтров достигается весьма любопытный эффект – увеличение скорости и повышение качества одновременно. Переход к целочисленным вычислениям позволяет существенно ускорить вычисления на микропроцессорах, которые используются в мобильных устройствах. А за счет комбинации подходов авторам удалось добиться скоростей распознавания, пригодных для применения в системах реального времени.

Необходимо отметить, что хотя принципы, лежащие в основе методов статистического обучения машин, достаточно изучены, зачастую не существует разумного способа объяснить действия уже обученной машины. В последнее время боль-

шое число работ предлагает способы обратной инженерии или хотя бы визуализации для глубоких нейронных сетей. В статье Б.М. Гаврикова и Н.В. Пестряковой предпринята успешная попытка исследования внутренних особенностей другого метода статистического обучения – так называемого метода полиномов.

На другом краю области распознавания образов лежат структурные, аналитические методы, не требующие больших обучающих выборок, но опирающиеся на наши априорные знания о распознаваемом объекте и законах оптики. В случае распознавания объектов трехмерного мира в видеопотоке основным требованием к алгоритмам распознавания является инвариантность к проективным преобразованиям, естественно возникающим при съемке. Точные в этом отношении алгоритмы давно известны для многоугольников и многогранников, но работа П.П. Николаева интересна тем, что в ней строго инвариантные алгоритмы распознавания строятся для гладких овалов, что оказывается гораздо сложнее, но при этом очевидно ближе к реальности.

В работе Т.В. Манжикова, О.А Славина и др. исследованы несколько способов применения *N*-грамм (на примере триграмм) для улучшения результатов распознавания кадров видеопоследовательности, полученной съемкой мобильным устройством паспорта гражданина РФ. Улучшение состояло в повышении точности распознавания при сохранении информативности оценки надежности. Были рассмотрены два алгоритма: один из них опирался на гипотезу зависимости символа от двух соседних символов, а второй был основан на вычислении маргинальных распределений с использованием графов на основе байесовских сетей. Результат позволяет заметно повысить точность распознавания отдельных полей.

Работа К.Б. Булатова, В.Ю. Кирсанова и др. посвящена проблемам, с которыми приходится столкнуться при решении задачи оптического распознавания текстовых строк. Авторы исследуют несколько простейших подходов к интеграции результатов распознавания в видеопотоке с анализом их особенностей и недостатков. Предложен алгоритм на основе выравнивания входных последовательностей при помощи модифицированного редакционного расстояния и проанализирован этот подход сравнительно с другими методами: предложенный метод показал улучшение точности распознавания в видеопотоке для трех из четырех испытуемых полей паспорта РФ.

Особенности использования мобильных устройств в качестве платформы для решения задач распознавания текстов на примере документов, удостоверяющих личность, рассматриваются в статье Д.П. Полевого, К.Б. Булатова и др. Авторам удалось уложить полный цикл получения, обработки и распознавания видеопотока непосредственно на устройство. Это позволяет наиболее полно воспользоваться преимуществами доступа к последовательности кадров, несмотря на высокую вариативность исходных объектов и характерный для слабо контролируемых условий съемок широкий спектр искажений. Ключевым аспектом работы является новая активная схема управления параметрами съемки и выбора зон обработки в процессе распознавания, что означает переход от традиционных пассивных схем ввода к системам с обратной связью.

Обычно задачу компьютерной трансмиссионной томографии относят либо к области интегральных преобразований, либо (в последнее время все чаще) к области численного решения больших систем линейных алгебраических уравнений. Однако физически исходные данные томографии представляют собой видеопоток рентгеновских проекций ригидного объекта. При использовании лабораторных источников излучения пренебрегать этим фактом уже нельзя, поскольку длительность съемки достигает нескольких часов, а установку нельзя считать идеально жесткой как минимум из-за теплового расширения. Работа М.В. Чукалиной, Д.П. Николаева, А.В. Бузмакова и др. посвящена созданию адекватной модели искажений, возникающих на установках подобного класса.

В работе Д.А. Девяткина, Р.Е. Суворова, И.А. Тихомирова, О.Г. Григорьева представлены результаты анализа основных критериев оцен-

ки научных проектов, используемых научными фондами, и предложен новый подход к определению значимости этих критериев, основанный на методах машинного обучения. Проведены экспериментальные исследования по определению значимости критериев оценки научных проектов на примере инициативных конкурсов РФФИ. Сделаны выводы о возможности использования предложенного подхода для верификации итоговых оценок экспертов, а также для проверки значимости вновь вводимых критериев.

Следует отметить, что полученные в проектах фундаментальные результаты во многих случаях легли в основу разрабатываемых прикладных систем, успешно реализованных или реализуемых в настоящее время. Алгоритмы распознавания, разработанные в проектах №№ 13-07-12170 и 13-07-12172, непосредственно применены в программно-аппаратном комплексе ввода документов, успешно внедряемом во многих организациях. Часть результатов, полученных при выполнении проекта № 13-07-12171, использована в промышленной разработке фотоаппарата нового типа для панорамной съемки. Алгоритмы, предложенные в проекте № 13-07-12106, применяются в разрабатываемой системе управления для беспилотного автомобиля. Примеры можно продолжить.

В то же время проведенные работы представляют собой первый шаг на пути построения теории и развития методов распознавания на мобильных устройствах. Именно теперь появилась возможность решить ряд принципиальных задач, связанных с интеграцией видеопотока не только на серверах, но и на мобильных устройствах. Это открывает огромные перспективы в развитии новых методов для таких дисциплин, как робототехника, управление движением летательных, наземных и подводных аппаратов, построение фототехники нового поколения и т.п.

Videostream. A New Approach to the Rigid Objects Recognition

S.V. Emelyanov

For the last 10-15 years, we see an unprecedentedly high interest in fundamental and applied research on image processing and recognition. It is associated primarily with the rapid development of digital photography, allowing to receive video-sequences of high-resolution photos on a relatively low-cost devices.

This has led to completely new approaches to the solution of all important problems of recognition, starting with the documents recognition and up to the human faces, natural scenes and ice conditions data recognition.

All over the world a lot of money is invested in this research area. The list of such foundations includes, on the one hand, such organizations as the National Science Foundation (NSF) or Japanese Institute for Research in Physics and Chemistry (RIKEN), which generally support the strictly scientific research, and on the other hand, corporations such as Google or Microsoft, which fund large goal-oriented projects of civic (such as an unmanned automobile) or military (for example, optical targeting systems) applications.

In 2013 the Russian Foundation for Basic Research (RFBR) has opened a three-year subject-oriented interdisciplinary research (ofi_m) "Perception and analysis of color images in the videostream and recognition of complex rigid objects". In the scope of projects being implemented under this theme a number of fundamental results had been achieved, all of them have theoretical and practical value. Original papers that describe the most of these results are presented in this themed issue.

Originally, the theme was limited to the recognition of so-called rigid objects, i.e. objects, which do not change their shape. For example, an automobile is a rigid object, and a horse is not. Although there are, of course, many intermediate cases. It is not always obvious how to classify a "curved" page of the open book or a tree in strong winds. Recognition of such objects in videostream has crucial advantages as opposed to their recognition on still pictures. Of course, it is always possible to recognize an arbitrary (random) frame or try to choose in some sense the best image from the input sequence. However, this is still rather tentative the first step.

If we have multiple frames containing the image of the same object, it is often possible to create a universal coordinate system on all the pictures. Then the whole sequence of input images could be viewed as a set of independent experiments, and a number of values in each coordinate point - as a realization of a random variable. Through this approach, methods of probability theory become applicable to the study of pattern recognition algorithms, without any reservations that usually accompany such a process.

In addition, most modern cameras have mechanisms to control the settings (focus, exposure, backlighting, white balance, etc.). This allows to include the feedbacks into the system and to solve the task of recognition using the apparatus of the control theory.

At the same time we must understand that during the videostream processing a number of technical problems appears. The first group of problems is due to uncontrolled shooting conditions (lighting, foreshortening, movement, etc.). Consequently, the working models must be robust against the image distortions. The source of the second group of problems is that the videostream is a huge array of information that must be processed in real time. This requires the creation of computationally efficient algorithms for both the proper recognition and processing of data. The third group of the problems is connected to the processing of videoframes captured with mobile devices, which generally have much lower resolution than images obtained with a photo camera or a scanner. The videoframe of a mobile device generally has a 2-4 megapixels resolution as opposed to photo cameras with resolutions of tens of megapixels.

Moreover, most of the practically important tasks of the videostream recognition must be solved on the mobile device itself. The transmission of the detailed information array to powerful servers is usually not possible for various reasons. Here are a few examples: car autopilot cannot spend time on data transfer; the satellite has no ability to transfer large amount of information due to the lack of a sufficiently wide

communication channel; autonomous underwater vehicles generally can only transmit a very limited amount of data.

Even in such relatively obvious task as passport recognition by means of a mobile device, there are serious reasons for performing the recognition on the device itself, because of legal restrictions on sensitive personal data processing and transfer. Moreover, the videostream is much harder to encrypt reliably in comparison with the recognition results which are measured not in tens of millions, but only in a few hundred bytes.

Within this issue of "RFBR Journal", the questions of a videostream recognition are discussed: usage of "deep learning" algorithms; creation and optimization of statistical methods of text recognition, which a resistant to distortions arising in the process of uncontrolled shooting; different approaches to the application of syntactic methods of pattern recognition, based on both the statistical properties of language and statistical models of videostream pattern recognition; the construction of recognition systems entirely based on the videostream; the creation of a distortions model in computed tomography; as well as analytical recognition methods based on projective invariants.

In the paper authored by T.S. Chernov, D.A. Ilin et al. two methods of printed text fields' segmentation are discussed. The first considered method develops classical approaches to text segmentation and comprises such stages as projection analysis, preliminary cutting and dynamic programming taking into account the probability scores of character recognition. The second method uses extensively the machine learning approaches, particularly convolutional and recurrent neural nets, this allows developing the segmentation algorithms without numerous heuristics methods tied to specific document field types and also increases the algorithms robustness to various distortions occurring while video-shooting by means of mobile devices.

The paper by E.G. Kuznetsova, I.V. Polyakov et al. considers the incremental learning approach to solving the representativeness problem of training data sets for machine learning. Authors described the structural model of unsupervised incremental learning process of the pattern recognition machine while a priory model of the observed scene evolution was used as a teacher. An option of extra-learning model for the task of moving objects detection in the videostream along with a set of necessary computer vision algorithms is suggested. Interestingly, the analogous cases, found by such procedure, are supposed to be corrected "on-the-fly" by means of the similar-cases-specialized search structure

E.E. Limonova, A.V. Sheshkus et al. investigated several methods of acceleration of images recognition by neural networks. The first method proposes to use the fixed-point arithmetic that allows to accelerate symbols recognition. The other two methods involve modification of the structure of a neural network convolutional layer in order to reduce the computational complexity. Through the use of a linear combination of separable filters quite an interesting effect is achieved – the increase in speed and quality simultaneously. The transition to the integer computing could significantly speed up the calculations on microprocessors, which are used in mobile devices. By means of the approaches combination, the authors were able to achieve speeds of recognition suitable for use in real-time systems.

It should be mentioned that while core principles of statistical machine learning are well studied, often there is no rational way to explain the properties of the already trained machine. Lately a large number of papers proposes the methods of reverse engineering or at least visualization for deep neural networks. In their paper B.M. Gavrikov and N.V. Pestryakova made a successful attempt to study the internal features of another method of statistical learning – the so-called polynomial regression method.

On the other side of pattern recognition area there are structure analysis methods which do not require very large training data sets but rather are based on our prior knowledge about the recognized object and rules of optics. Applied to the problem of recognition of three-dimensional objects in videostream, the main requirement for recognition algorithms is resistance to projective transformations naturally arising when shooting. Precise algorithms are known for such objects as polygons and polyhedrons, but the P.P. Nikolayev study is interesting because he constructed the precise invariant algorithms for smooth ovals, which proves to be more difficult but obviously closer to reality.

T.V. Manzhikov, O.A. Slavin et al. investigated the N-grams usage for the correction of the recognition results of images documents through the example of text fields of a passport of the Russian Federation citizen. For the trigrams, the two algorithms for adjusting the recognition results are suggested. One of the algorithms is based on the use of trigram probabilities combined with recognition estimates, which are also interpreted as probabilities. The second algorithm is built on the definition of marginal distributions and the calculations on graphs based on the Bayesian networks. The algorithms usage could significantly improve the recognition accuracy of individual text fields.

The paper authored by K.B. Bulatov, V.Y. Kirsanov et al. is devoted to the problem of an integration of the results of the text fields identification in the videostream received via a mobile device camera. The authors offered the algorithm based on the alignment of the input strings by means of a modified edit distance and performed the comparative analysis of the algorithms' outputs. The proposed method has shown the improvement of recognition accuracy in videostreams for three of the four text fields of the Russian Federation citizen passport.

In their paper D.V. Polevoy, K.B. Bulatov et al. discussed some features of mobile devices usage as a platform for solving the tasks of text recognition through the examples of identity documents. The authors managed to pack a complete cycle of receiving, processing, and recognition of the videostream directly in the device. This allows the authors to use the advantage of the access to frames sequence despite the high variability of the original objects and a wide range of distortions characteristic for poorly controlled shooting conditions. The key aspect of this work is the new active control scheme for the shooting settings and selections of the treatment areas during the recognition process that means the change from the traditional passive circuit input to the feedback systems.

Usually, researchers attribute the task of computer transmission tomography either to the field of integral transforms, or (more often in recent times) to the field of numerical solution of large systems of linear algebraic equations. However, physically the original data input for tomography is a videostream of X-ray projections of a rigid object. While using the laboratory radiation sources one must not neglect this fact, since the shooting duration is up to several hours, and the equipment cannot be considered perfectly rigid at least due to thermal expansion. The paper by M.V. Chukalina, D.P. Nikolaev, A.V. Buzmakov et al. is devoted to the creation of an adequate distortions model for this type of setups.

In the paper authored by D.A. Devyatkin, R.E. Suvorov, I.A. Tikhomirov and O.G. Grigoriev the result of analysis of funding criteria of scientific projects, commonly used by funding organizations, are presented, and the novel approach for estimating importance of funding criteria, used to assess scientific projects, is offered. The authors experimentally determined the relevance of scientific criteria for projects estimation through the examples of RFBR grant competitions. The authors concluded that the suggested approach can be applied for the verification of experts' final estimates as well as for the check of the importance of the newly introduced criteria.

It should be noted that the projects' fundamental results in many cases have formed the basis for the development of applied systems, successfully implemented or being implemented at present time. The recognition algorithms developed in the projects Nos. 13-07-12170 and 13-07-12172 are directly applied in the software-and-hardware system for documents entry, which is successfully implemented in many organizations. Some results obtained in the project No. 13-07-12171 were used in the commercial development of a new-generation cameras for panoramic photography. The algorithms proposed in project No. 13-07-12106 are used to develop the control system for an unmanned automobile. The list of examples could be continued.

At the same time the conducted studies are in a sense the first step to the creation of a common theory and to the development of the recognition techniques on mobile devices. It is now possible to solve a number of principled problems, associated with the videostream integration, not only on high-performance servers but also on mobile devices. This opens up great prospects for the development of new techniques for some disciplines such as robotics, motion control of aircraft, surface and underwater vehicles, creation of the new-generation photographic equipment, etc.

Разработка алгоритмов технического зрения для обеспечения инкрементного неконтролируемого обучения в задачах детектирования образов движущихся сложноструктурированных объектов*

Е.Г. Кузнецова, И.В. Поляков, Д.П. Николаев, Д.Н. Мацнев

В настоящей статье рассматривается проблема репрезентативности данных при обучении распознающих машин в индустриальных системах технического зрения и подход к ее решению на основе инкрементного обучения. Описывается структурная модель процесса неконтролируемого дообучения распознающей машины, использующей в качестве учителя априорную модель эволюции наблюдаемой сцены. Предлагается вариант модели дообучения для задачи детектирования движущихся объектов в видеопотоке, а также набор необходимых алгоритмов машинного зрения. В качестве обучаемого детектора движущихся объектов используется древовидная модификация алгоритма Виолы–Джонса, причем предлагается ряд техник для повышения ее качества при обучении и дообучении. В качестве модели эволюции рассматривается модель поступательного движения. Для ее использования предлагается новый метод прослеживания образов на основе комбинации оценок оптического потока и сопоставления ортотропных границ на изображениях. Приводятся результаты вычислительных экспериментов, демонстрирующие надежность предложенных подходов.

Ключевые слова: обучение машин, неконтролируемое обучение, инкрементное обучение, Виола–Джонс, прослеживание объектов.

** Работа выполнена при финансовой поддержке РФФИ (проект № 13-01-12106).*

Введение

В настоящей работе рассматривается адаптация модели неконтролируемого инкрементного обучения классификатора для распознавания изображений в режиме реального времени (проект РФФИ № 13-01-12106-офи_м «Методы обучения и дообучения для систем видеоклассификации сложноструктурированных объектов в неконтролируемых условиях») к решению задачи детектирования движущихся объектов.

Методы технического зрения на основе контролируемого (supervised) обучения машин, признанные наиболее эффективным решением широкого спектра лабораторных задач, активно находят свое применение в промышленных и военных системах технического зрения для детектирования и классификации объектов. Однако высокая сложность и вариативность объектов реального мира и неконтролируемость условий наблюдения существенно затрудняют их внедрение. Проблематичность обучения машин для реальных задач обусловливается сложностью построения репрезентативной и сбалансированной для совокупности реальных условий обучающей выборки маркированных данных ввиду

КУЗНЕЦОВА
Елена Геннадьевна
Институт проблем передачи информации им. А.А. Харкевича РАН

ПОЛЯКОВ
Игорь Викторович
Институт проблем передачи информации им. А.А. Харкевича РАН

НИКОЛАЕВ
Дмитрий Петрович
Институт проблем передачи информации им. А.А. Харкевича РАН

МАЦНЕВ
Дмитрий Николаевич
Институт проблем передачи информации им. А.А. Харкевича РАН

высокой сложности и вариабельности распознаваемых образов и фоновых сцен, следующей как из сложности природы реальных объектов, так и из искажений в результате динамических изменений фона регистрируемой сцены (например изменения настроек экспозиции регистрирующей камеры, наличия фонового движения) и изменения положения объекта в пределах наблюдаемой сцены.

Обычным подходом к повышению репрезентативности данных для обучения является расширение обучающей выборки новыми образцами распознаваемых объектов с последующим переобучением машины. Получение новых образцов может осуществляться путем как маркировки оператором естественных (то есть полученных регистрирующей камерой в процессе эксплуатации системы) изображений, так и аугментации (data synthesis) промаркированных ранее данных для обучения.

Аугментация данных представляет собой построение новых входных образцов для обучения машины путем применения различных преобразований к естественным изображениям (локальные и глобальные аффинные преобразования, варьирования яркости и контрастности) [1]. Применение аугментации данных позволяет достичь заметного повышения эффективности обучаемых машин и может решать проблему сбалансированности обучающих данных, однако в реальных задачах распознавания образов возможности этого подхода существенно ограничены невозможностью качественного моделирования сложных искажений образов реальных трехмерных объектов на их двумерных изображениях (например проблематично смоделировать тени, падающие на объемные объекты) и невозможностью добавления новых образцов, прообразы которых существенно отличаются от имеющихся в исходной выборке. Поэтому для реальных задач аугментацию данных целесообразно использовать только в сочетании с методами расширения выборки естественными изображениями.

Простейший и широко распространенный на практике подход к расширению выборки естественными изображениями заключается в добавлении выбираемых случайно и маркируемых оператором изображений, полученных в процессе эксплуатации системы. Однако случайный выбор добавляемых в выборку образцов не решает проблемы сбалансированности данных для обучения (неравномерности распределения данных между классами и внутри классов относительно различных условий наблюдения, imbalanced data learning), что приводит к существенному снижению качества машин, обучаемых с использованием большинства стандартных алгоритмов, так как они предполагают обучение на сбалансированных данных с равной стоимостью ошибки распознавания для всех образцов [2]. Известны две основные группы подходов к решению этой проблемы: 1) построение сбалансированной выборки на основе исходной (случайное удаление/добавление элементов выборки, итеративное обучение машин на динамически формируемых подмножествах [3], предварительная кластеризация данных для обучения с выравниванием объемом кластеров и удалением выбросов [4, 5]); 2) модификация алгоритмов обучения на несбалансированных данных (модификация матрицы стоимости ошибки классификации для различных классов/кластеров несбалансированной выборки [6]). Общей проблемой всех этих подходов можно считать высокие временные и вычислительные затраты, связанные с необходимостью получения, хранения и ручной маркировки избыточных объемов данных, которые фильтруются в процессе обучения, и переобучением машин на разрастающихся выборках.

Известным подходом к решению проблемы необходимости обучения на избыточных объемах маркируемых оператором данных является разработка модулей, реализующих полуконтролируемое (semisupervised) и неконтролируемое (unsupervised) дообучение машины путем формирования и автоматической маркировки данных для дополнительного обучения (incremental learning) в процессе эксплуатации системы.

В видеосистемах распознавания движущихся объектов наибольшее распространение в рамках этого подхода получили методы на основе валидации срабатываний машины алгоритмами, реализующими проверку ограничений на параметры движения объекта в рамках заданной модели движения. В качестве таких алгоритмов используются методы прослеживания объектов

[7] либо методы сегментации движения [8]. Осуществляется оценка достоверности срабатываний машины на основе алгоритмов оценки движения, вероятные ошибки автоматически маркируются с помощью этих алгоритмов и добавляются в выборку для дообучения машины. Получаемая таким образом выборка не требует маркировки оператором и в меньшей степени избыточна (в сравнении со случайным выбором образцов), так как формируется из множества потенциальных ошибок распознающей машины.

Однако качество получаемой таким образом дополнительной выборки существенно зависит от надежности работы алгоритмов оценки движения. Поэтому при разработке методов в рамках этого подхода одним из наиболее существенных вопросов является выбор надежного алгоритма оценки движения, определяющего качество набора и маркировки обучающих данных. Решение этого вопроса усложняется требованием к работе алгоритма в режиме реального времени в процессе эксплуатации системы в условиях ограниченной мощности индустриальных вычислителей.

В наиболее известном в рамках концепции инкрементного дообучения с привлечением модели движения подходе Tracking–Learning–Detection [7] для детектирования и долгосрочного прослеживания движущихся объектов в качестве алгоритма оценки движения предлагается метод прослеживания объектов на основе оценок оптического потока Median-Flow tracker, однако отмечается целесообразность дальнейшего улучшения этого метода. При этом в работе [9] отмечается, что трудность реализации метода для работы в реальном времени ввиду высокой вычислительной сложности является характерной проблемой для наиболее надежных современных методов прослеживания движения объектов. Поэтому формируемая в реальном времени с помощью методов прослеживания движения обучающая выборка, как правило, требует дополнительной ручной проверки для фильтрации неизбежных ошибок алгоритма прослеживания и распознающей машины или модификации алгоритмов обучения для снижения зависимости качества обучения машины от «замусоренности» выборки.

Данная работа посвящена разработке модуля неконтролируемого инкрементного обучения машины для детектирования движущихся объектов с использованием метода прослеживания движения и дополнительной автоматической фильтрацией его ошибок в формируемой выборке путем кластеризации образцов для обучения и проверки согласованности меток для кластеров схожих объектов. В качестве метода прослеживания предлагается устойчивый алгоритм определения параметров движения детектируемого объекта на основе комбинации разреженных оценок оптического потока и сопоставления краевых границ на изображении. Высокие скорость и надежность работы алгоритма достигаются за счет использования ограничений на параметры движения объекта в рамках кулисной модели сцены. В качестве обучаемой машины предлагается древовидная модификация алгоритма детектирования объектов Виолы–Джонса, обеспечивающая сопоставимую с классическим алгоритмом производительность, но более низкую вычислительную сложность инкрементного обучения. Приводятся результаты вычислительных экспериментов, демонстрирующих быстродействие и качество предложенных подходов.

Модель модуля инкрементного обучения

Предлагаемая модель модуля дополнительного обучения для классификации образов содержит статический обучаемый классификатор, дополненный высокоуровневой моделью эволюции распознаваемого объекта в наблюдаемой сцене, которая обеспечивает автоматическое формирование маркированной обучающей выборки для неконтролируемого дообучения классификатора в режиме реального времени *(рис. 1)*.

Формирование данных для дообучения классификатора осуществляется путем кэширования его ошибок, детектируемых в реальном времени путем валидации ответов детектора высокоуровневой моделью распознаваемого объекта. Алгоритм кэширования использует метод иерархической кластеризации FLANN [10], обеспечивающий быстрый ответ на запросы на поиск как ближайшего по метрике, так и всех «похожих» объектов. При-

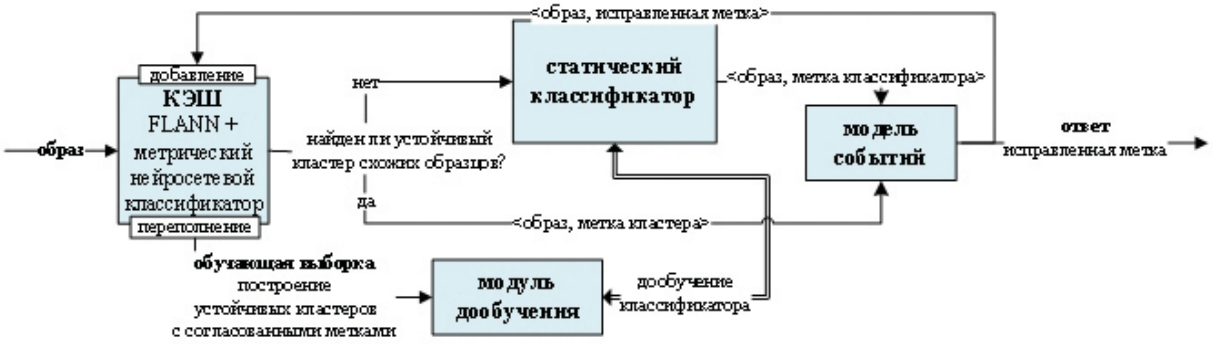


Рис. 1. Схема модуля неконтролируемого обучения.

менение кластерной структуры позволяет парировать известные ошибки статического детектора до момента дообучения (если для образца сформировался значительный по мощности кластер близких к нему объектов с согласованными метками модели, образец не подается на вход детектора, а в качестве ответа распознающего модуля используется метка кластера). По мере заполнения дерева поиск замедляется, одновременно с этим сохраненная в дереве дообучающая выборка становится более консистентной. При «переполнении» кэша проводится инкрементное обучение статического детектора. При этом кластерная структура данных обеспечивает возможность дополнительной проверки согласованности меток для кластеров схожих изображений, что обеспечивает высокую робастность к возникновению ошибок уровня модели.

Обучаемый детектор объектов

В рамках адаптации описанной ранее структурной модели к задаче детектирования движущихся объектов в качестве статического обучаемого детектора объектов предлагается использование предложенной [11] обобщающей модификации алгоритма Виолы–Джонса.

Классический детектор Виолы–Джонса *(рис. 2a)* представляет собой последовательность (каскад) сильных классификаторов, которые строятся на основе алгоритма AdaBoost [12] как линейная комбинация слабых классификаторов, основанных на пороговой проверке хааароподобных признаков [13], и используется для проверки гипотезы о наличии искомого объекта в регио-

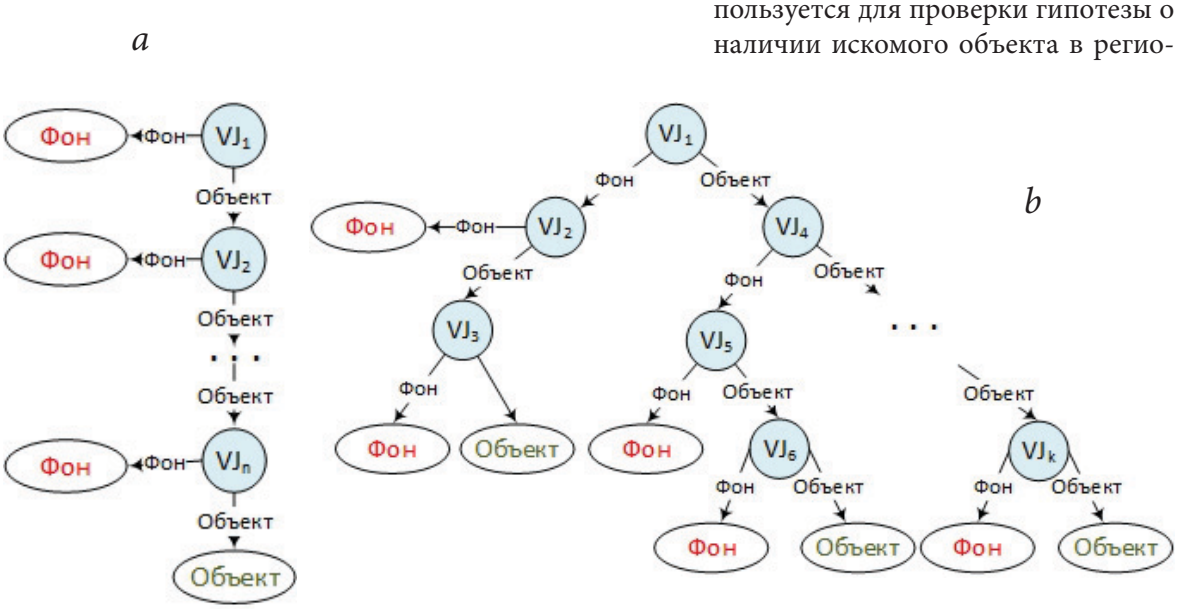


Рис. 2. Структуры машин на основе алгоритма Виолы–Джосна: a - каскад, b – древовидная модификация.

DOI: 10.22204/2410-4639-2016-092-04-17-37 № 4 (92) октябрь–декабрь 2016 г.

нах изображения, взятых с различными сдвигами и масштабами.

Предлагаемая модификация является обобщением каскадной структуры сильных классификаторов в виде решающего дерева *(рис. 2b)*. Преимуществами обобщающей модификации в сравнении с классическим детектором Виолы–Джонса является меньший разрыв времени работы в худшем и среднем случае, достигаемый за счет сбалансированной структуры дерева, что важно для систем реального времени в условиях ограничений на время обработки каждого входного изображения, и возможность инкрементного обучения на новых данных без сохранения исходной обучающей выборки путем добавления новых уровней к листам дерева, что решает проблему хранения разрастающихся данных при реализации модуля инкрементного обучения и сокращает время инкрементного обучения, обеспечивая возможность его работы в режиме реального времени.

Проверить гипотезу о том, что предложенная модификация действительно обеспечивает более высокую скорость распознавания в худшем случае, можно только экспериментально на той или иной обучающей выборке. Основным параметром выборки, помимо объема, является визуальное разнообразие классифицируемых (детектируемых) объектов. Известно, что многие усовершенствования классических эффектов хорошо показывают себя на «простых» задачах, но не дают улучшения в реальных условиях. Поэтому в эксперименте использовались данные, полученные в реальных условиях *(рис. 3)* видеосистемой автоматической классификации транспортных средств (АКТС) [14]. Была обучена пара машин (классический детектор Виолы–Джонса и обобщающая модификация [11]) для решения задачи обнаружения колес транспортных средств (ТС) *(рис. 3)*. Для обучения и оценки производительности использовался набор вручную маркированных кадров (~12 тыс. изображений, содержащих ~3000 изображений колес транспортных средств).

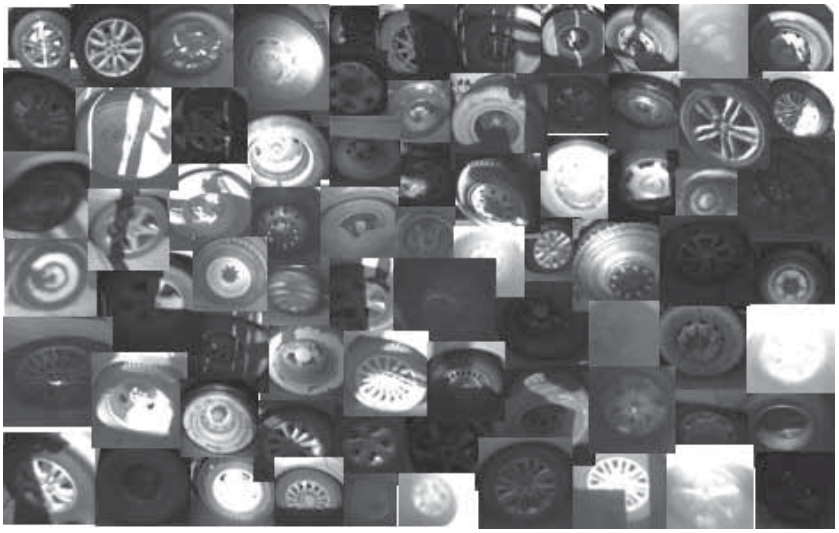


Рис. 3. *Образцы из тестового набора детектируемых образов.*

Результаты вычислительных экспериментов подтвердили преимущества древовидного классификатора по быстродействию в худшем случае на ~6.1%, а в лучшем случае – ~0.96% при заметном улучшении показателей качества классификации (в 1.1 раза по положительным примерам, в 2.2 раза – по отрицательным).

Начальное обучение детектора объектов. Набор данных для первоначального обучения детектора образов в режиме оффлайн является нерепрезентативным относительно реальных условий, что и обусловливает низкое качество работы начального детектора и необходимость его инкрементного дообучения. Однако качество начального обучения существенно влияет на результативность процесса инкрементного дообучения, так как ответы детектора используются для инициализации алгоритма прослеживания.

Для повышения качества работы начального детектора предлагается использование упомянутой ранее техники аугментации данных для обучения – одного из известных подходов к решению проблемы обучения машин на несбалансированных обучающих выборках. Целью аугментации данных является устранение несбалансированности обучающей выборки и повышение вариативности входных признаков машины, используемых при обучении.

Для исследования эффективности применения техники аугментации данных для начального обучения детектора объектов в разрабатываемом модуле

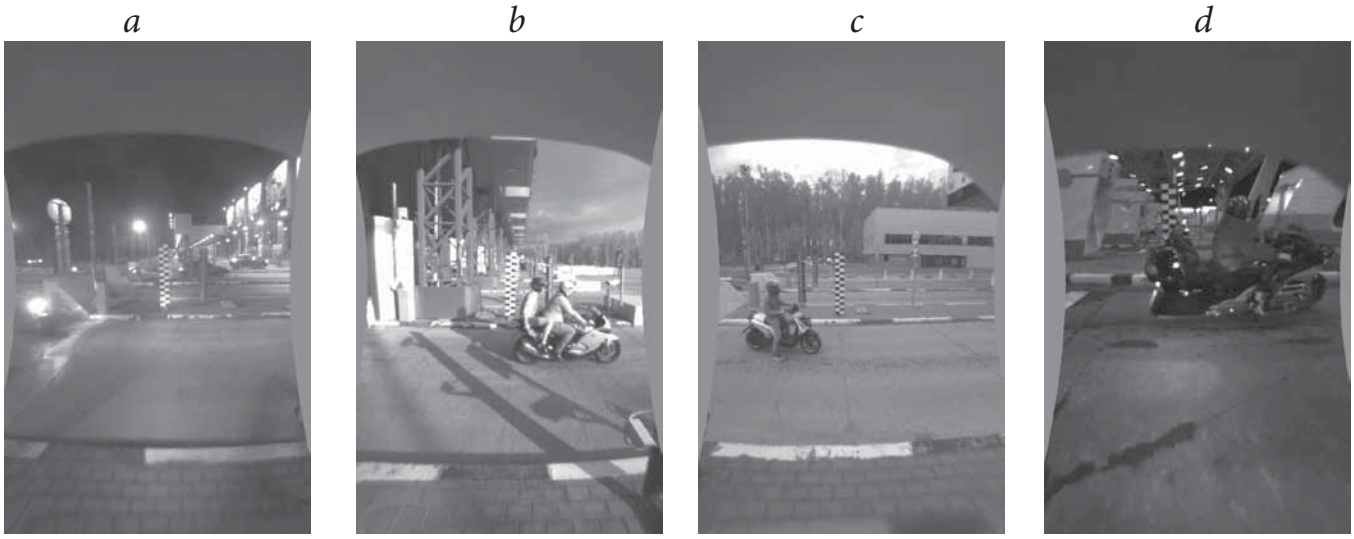


Рис. 4. *Примеры кадров АКТС с изображениями мотоциклов (a-d).*

был проведен ряд вычислительных экспериментов по обучению машин для детектирования мотоциклов на изображениях, регистрируемых видеокамерами АКТС *(рис. 4a-d)*. Обучающая выборка состояла из ~15.5 тыс. положительных образцов *(рис. 4a-d)*, полученных в режиме реального времени с полос АКТС.

Были использованы следующие методы аугментации данных, моделирующие характерные для реальных сцен искажения детектируемых объектов:

1. Модификация шкалы яркости изображения для моделирования искажений яркости и контрастности в условиях неравномерного освещения и автоматически изменяющихся в реальном времени настроек экспозиции регистрирующей камеры.

Пусть I – исходное изображение с 8-битными значениями в диапазоне [0, 255]. Тогда результат модификации его яркости J определяется в каждой точке $(x, y) \in Q$ где Q – пространственный домен изображений, определяется как

$$i(x, y) = LUT[i(x,y)], \quad (1)$$

где $LUT = [I_0, I_1, ..., I_{255}]$, $0 \le I_i \le 255$ $\forall i \in [0, 255]$ – таблица подстановок (lookup table), генерируемая динамически как псевдослучайная монотонная гладкая дискретная функция, принимающая значения от 0 до 255 согласно следующей схеме:

$$1. \begin{cases} I_t = 0, i = 0 \\ I_{t+1} = I_t + \frac{rand_{0...1}}{K}, \forall i \in [0,255], \end{cases}$$

где $rand_{0...1}$ – псевдослучайно (нормальное распределение) сгенерированное в диапазоне [0, 1] число, K – параметр, постоянная величина, регулирующая гладкость случайно генерируемой функции.

2. $I_i = 255 \cdot I_i / I_{255}$, $\forall i \in [0, 255]$.

На *рисунке 5b*, с изображен результат случайного искажения яркости *изображения 5a*. Соответствующие шкалы яркости представлены на *рис. 5d–f*.

2. Моделирование геометрических искажений. Детектор объектов осуществляет проверку гипотезы о наличии искомого объекта для всевозможных регионов входного изображения, имеющих различные размеры и координаты на изображении. Перебор положений (x, y), $x \in W$, $y \in H$ (где (W, H) – размер изображения) и размеров $s \in [s_{min}, s_{max}]$, $s_{min} < s_{max} \le W \cdot H$ регионов осуществляется с дискретными шагами dx, dy, ds соответственно. Следствием дискретизации перебора регионов, подаваемых на вход детектора образов, являются неточности выбора регионов, содержащих искомые объекты, что приводит к снижению качества работы детектора, так как обычно его обучение проводится на регионах, соответствующих точным окаймляющим прямоугольникам объектов, промаркированных вручную.

Для повышения устойчивости обучаемой машины к неточностям выбора региона предлагается использовать аугментацию обучающей выборки в виде генерирования новых положительных примеров для обучения путем случайных малых сдвигов вертикальных и горизонтальных границ окаймляющего прямоугольника маркированного вручную объекта *(рис. 6)*.

DOI: 10.22204/2410-4639-2016-092-04-17-37 № 4 (92) октябрь–декабрь 2016 г.

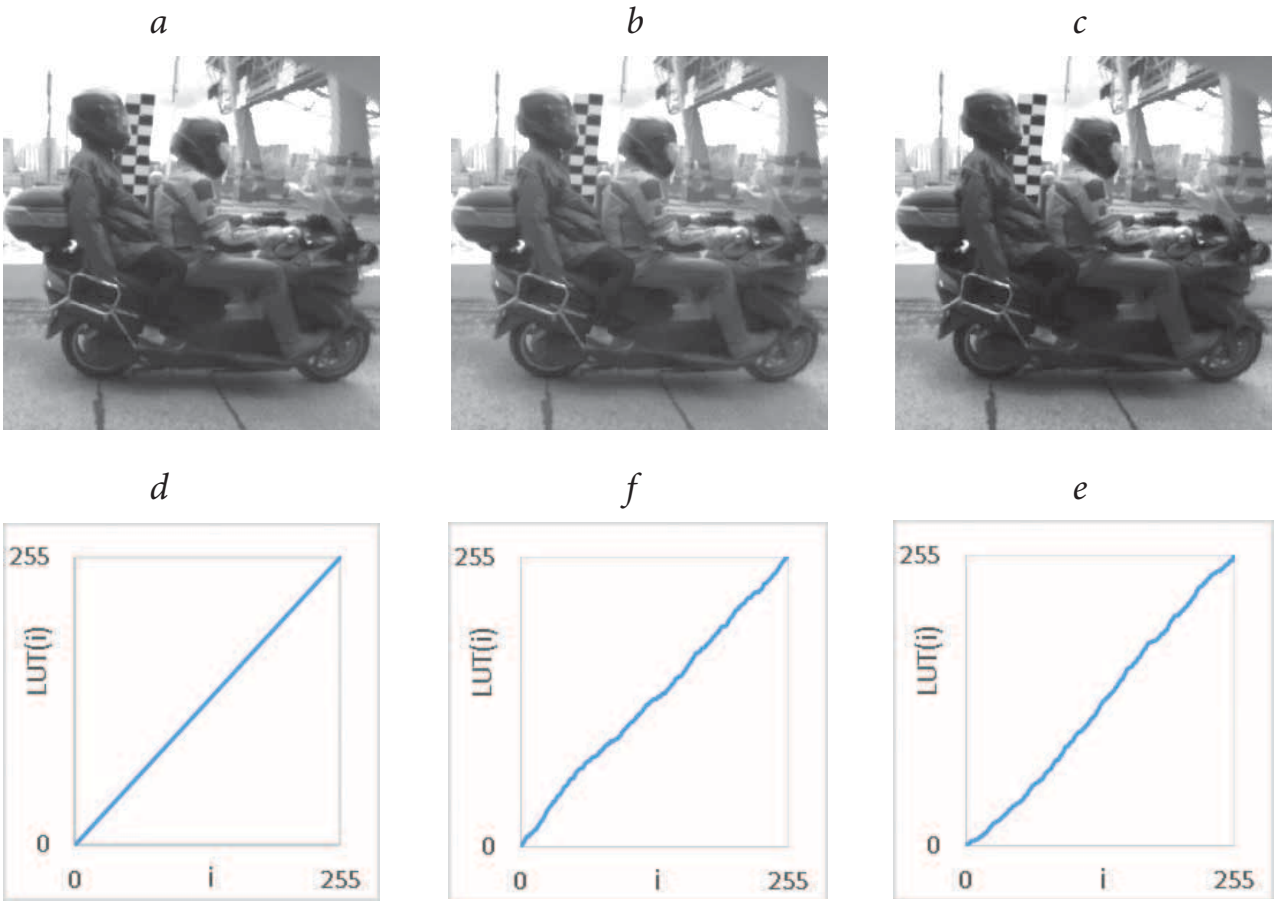


Рис. 5. *Примеры аугментации изображений с модификацией шкалы яркости: a – исходное изображение, b,c – модификации, d-f – соответствующие шкалы яркости.*

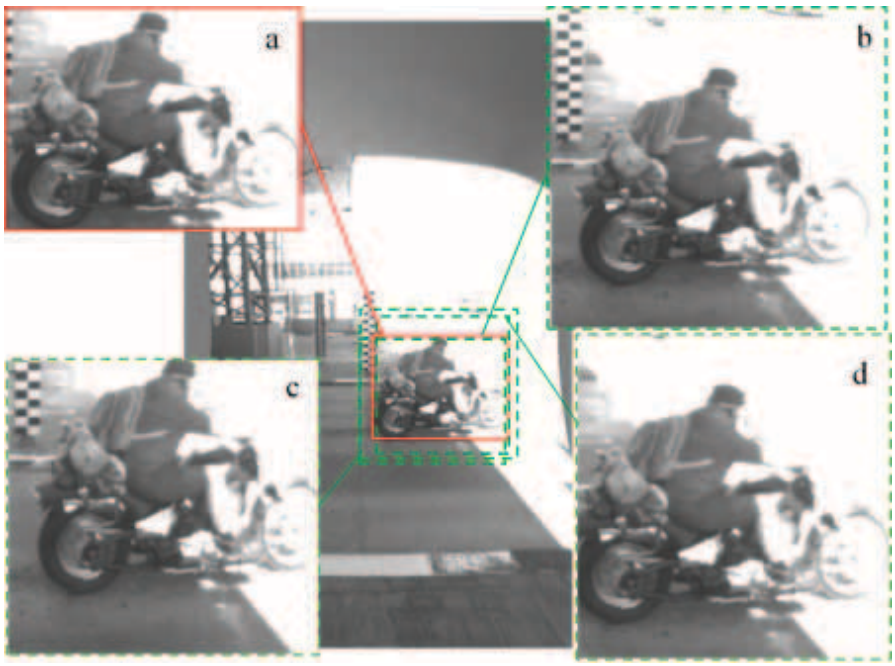


Рис. 6. *Примеры аугментации изображений со смещением линий границ окаймляющего прямоугольника: a – исходный регион объекта, b-d – результаты аугментации.*

Для исследования результативности применения аугментации данных был проведен вычислительный эксперимент по обучению пары детекторов Виолы–Джонса для детектирования мотоциклов на изображениях *(табл. 1)*. При тестировании полученных детекторов использовались следующие параметры перебора регионов на входных изображениях (2): $dx = 1$, $dy = 4$, $s_{min} = 1.2$, $s_{max} = 3$, $ds = 1.2$ для кадров видеоряда, предварительно сжатых в три раза. Случайный сдвиг сторон рамки окаймляющего прямоугольника при аугментации осуществлялся в пределах ±10% от ее измерений. Результаты экспериментов приведены в *таблице 1* (N_{wc} – число слабых классификаторов).

Обучение каждого классификатора проводилось до тех пор, пока ошибка второго рода не достигнет 0.01 (срабатывание на объект в одном случае из ста). При этом измеряли показатели ошибки первого рода (доля пропусков) и интегральный показатель времени работы (математическое ожидание числа сработавших слабых классификаторов на один запуск).

Как видно из *таблицы 1*, использование аугментации положительных примеров в обучающей выборке позволяет ускорить сходимость обучаемой машины к оптимальному решению (девять уровней детектора с аугментацией, 16 – без аугментации) при сопоставимом среднем времени работы и уменьшить ошибку первого рода в 0.1645/0.0860 ≈ 1.9 раза при сопоставимой ошибке второго рода, что демонстрирует существенное повышение устойчивости обучаемой машины к переобучению.

Метод прослеживания движущихся объектов в видеопотоке

В качестве модели событий, используемой в модуле инкрементного дообучения детектора, для отслеживания движущихся объектов в видеосистемах предлагается использовать ограничения на кинематику движения, реализуемые в виде модуля прослеживания объектов на основе оценки оптического потока.

Оптический поток представляет собой распределение видимых скоростей движения яркости элементов изображения, возникающий в результате относительного движения между наблюдателем и наблюдаемой сценой (*). На практике вычисление оптиче-

Таблица 1. *Результаты экспериментов по обучению детекторов мотоциклов*

Обучение без аугментации				Обучение с аугментацией			
Уровень дерева	Ошибка второго рода	Ошибка первого рода	N_{wc}	Уровень дерева	Ошибка второго рода	Ошибка первого рода	N_{wc}
1	0.0210	0.3721320	6	1	0.0110	0.406106	7
2	0.0355	0.2280660	9	2	0.0115	0.343031	7
3	0.0590	0.1193880	15	3	0.0225	0.18222	19
4	0.0645	0.1002590	10	4	0.0360	0.118939	20
5	0.0725	0.0958354	12	5	0.0430	0.0716792	24
6	0.086	0.0909151	19	6	0.0610	0.048532	19
7	0.0965	0.0754331	27	7	0.0715	0.0309049	24
8	0.1000	0.0798331	16	8	0.0792	0.0204564	21
9	0.1335	0.0408132	20	9	0.0860	0.0114894	21
10	0.140	0.0360354	23	-	-	-	-
11	0.1415	0.0339197	10	-	-	-	-
12	0.1470	0.0280738	8	-	-	-	-
13	0.1510	0.0232482	11	-	-	-	-
14	0.1605	0.0179654	22	-	-	-	-
15	0.1605	0.0143202	4	-	-	-	-
16	0.1645	0.0136899	21	-	-	-	-

DOI: 10.22204/2410-4639-2016-092-04-17-37 № 4 (92) октябрь–декабрь 2016 г.

ского потока применяется для восстановления проекции на плоскость изображения фактического движения элементов в пространстве наблюдаемой сцены, называемого полем движения. Обзор современного состояния разработок в области оценки оптического потока приведен в работе [9].

Определение (*) оптического потока предполагает постоянство интенсивности перемещающихся пикселей изображения во время движения. Пусть дана последовательность изображений видеоряда I: $Q\ p \rightarrow R$, где $Q \in R^2$ – домен изображения, T – дискретный интервал времени для изображений последовательности. Тогда уравнение сохранение яркости определяется следующим образом:

$$\frac{\partial I}{\partial t}(p(t), t) = 0. \qquad (3)$$

Дискретная аппроксимация (3) в данной точке $p \in Q$ в момент времени t дает уравнение (4):

$$I(p + w(p), t + 1) - I(p, t) = 0, \qquad (4)$$

где $w(p) = [u(p)\ v(p)]^T$ – оценка оптического потока в точке $p = [x\ y]^T$.

На практике, как правило, используется линеаризация (4) вида (5):

$$\nabla I(p) \cdot w(p) + I_t(p) = 0, \qquad (5)$$

где $\nabla = [\partial/\partial x\ \partial/\partial y]^T$ – оператор пространственного градиента, $I_t = \partial I/\partial t$ – частичная производная по времени.

Единственного линеаризованного уравнения сохранения яркости (3) недостаточно для восстановления двух неизвестных компонент $w(p)$. Эта неопределенность известна как проблема апертуры, предполагающая, что движение имеет линейную структуру, так как уравнение (3) не имеет однозначного решения, если не учитывается информация о соседних пикселах. Поэтому задача оценки оптического потока, как правило, представляется в виде задачи оптимизации некоторого функционала ошибки для локальных регионов изображения в рамках выбранной параметрической модели движения либо как задача глобальной оптимизации с регуляризирующим функционалом.

Необходимость итерационного решения задачи оптимизации неизбежно порождает проблемы применимости таких подходов в системах реального времени из-за высокой вычислительной сложности. Однако достижение требуемой скорости без снижения качества может быть достигнуто путем внесения более жестких ограничений на модель движения.

Предлагается быстрая реализация оценки поля движения в рамках так называемой «кулисной модели сцены»:

- все объекты являются твердыми телами и движутся с ограниченной скоростью по горизонтальной поверхности преимущественно перпендикулярно оси регистрирующей камеры;
- фоновое движение преимущественно отсутствует.

Область применимости этой модели – системы распознавания, в которой наблюдаемые объекты движутся в заданном направлении по горизонтальной поверхности. Это характерно для пропускных пунктов людей или транспорта, конвейеров с движущимися объектами и т.п.

Модель движения в рамках кулисной модели определяется следующим образом:

$$\begin{cases} v(p) = 0 \\ u(p) = \begin{cases} u_{bg}(p) = 0, p \in \text{фон} \\ u_{fg}(p) = 0, p \in \text{объект} \end{cases} \end{cases} \qquad (6)$$

Уравнение (6) с одной неизвестной величиной $u(p)$ может быть решено в каждой точке $p \in Q$ независимо. Это позволяет избежать вычислительных затрат, связанных с решением задачи оптимизации на регионах изображения, но порождает ошибки оценки оптического потока на отдельных пикселах, так как при прямом решении уравнения (6) теряются ограничения на гладкость поля оптического потока для окрестностей пикселов, поддерживаемые оптимизационными схемами.

Однако горизонтальная компонента оптического потока $u(p)$ согласно модели (6) может принимать лишь одно из двух значений $u(p)$ с $\{0, u_{fg}(p)\}$, что в предположении о независимости ошибок определения оптического потока в отдельных точках позволяет надежно определять параметры поля движения объекта как ненулевой локальный максимум на гистограмме распределения параметра u оптического потока (или принять гипотезу об отсутствии объекта в поле зрения камеры в случае отсутствия

выраженного ненулевого локального максимума), оцененного в каждой точке изображения p, где оптический поток определен, то есть:

$$\frac{\partial I}{\partial x} > \varepsilon, \qquad (7)$$

где ε – малая величина, параметр метода.

На *рисунке 7* изображены гистограммы оценок оптического потока для пары последовательных кадров, содержащих изображение частей движущегося ТС *(рис. 7a–c)*, и кадров, не содержащих движения *(рис. 7d–f)*. Как видно из *рисунка 7c*, на двумерной гистограмме $\partial I/\partial x(\partial I/\partial t)$ наблюдается два выдающихся локальных пика в $\partial I/\partial t = 0$, что соответствует нулевому смещению, или фону, и $\partial I/\partial x \,/\, \partial I/\partial t = u_{fg}$, что отвечает смещению наблюдаемого движущегося объекта. На *рисунке 8* продемонстрирован результат определения параметра u_{fg} поля движения объекта на последовательности кадров проезда ТС. *Рисунок 8m* получен как усредненное изображение всех кадров проезда ТС с учетом сдвигов, соответствующих смещению u_{fg}, найденному с помощью описанного алгоритма для каждого кадра видеопоследовательности проезда ТС.

В работе [15] представлен вычислительный эксперимент, демонстрирующий сравнение производительности оценки смещения на основе оптического потока, метода полного перебора всевозможных целочисленных смещений с минимизацией отклонений яркости и метода на основе сопоставления особых точек ASIFT с алгоритмом RANSAC с проективной моделью движения на примере видеоданных, полученных с полос АКТС. Эксперимент демонстрирует выигрыш производительности первого метода в сравнении со вторым и существенный проигрыш в производительности по сравнению с третьим. Однако следует отметить, что в эксперименте [15] правильным ответом (ground truth) считался достаточно свободный диапазон скоростей, так как части движущихся ТС неизбежно подвергаются аффинным искажениям; алгоритм ASIFT + RANSAC имеет несравнимо большую вычислительную сложность как собственно вычисления особых точек изображения ASIFT, так и поиска оптимума алгоритмом оптимизации RANSAC.

Отметим, что уравнение сохранения яркости (3) действительно только при условии, что интенсивность перемещающихся пикселей остается постоянной во время движения, однако в видеоизображениях реального мира это условие может нарушаться в результате изменений освещения. Большей устойчивостью к варьированиям яркости характеризуются алгоритмы оценки движения на основе сопоставления характерных признаков (текстурных и цветовых признаков, ключевых точек [16]). Однако их характерными недостатками являются их более низкая точность в сравнении с алгоритмами на основе оценок оптического потока и неопределенность при об-

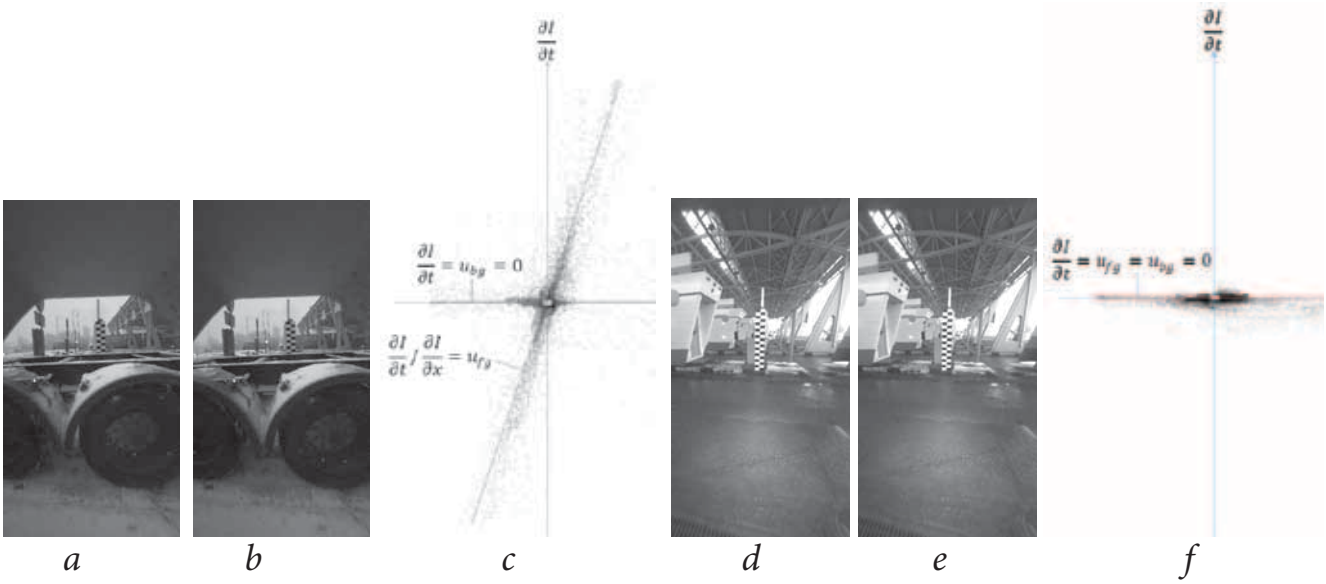


Рис. 7. *Примеры гистограмм оценок оптического потока $\partial I/\partial t(\partial I/\partial x)$ для изображений I в момент времени t (для наглядности значения точек на гистограммах дважды прологарифмированы): a-c – объект движется в рамках кулисной модели сцены; d-f – движение отсутствует (a, d – I(t); b, e – I(t+1); c, f – $\partial I/\partial t(\partial I/\partial x)$).*

Рис. 8. Визуализация работы алгоритма на основе оценок оптического потока для определения параметра поля движения движущегося ТС в видеопоследовательности: a-f – исходные изображения видеоряда (приведен каждый сороковой кадр); g-l – гистограммы оценок оптического потока, соответствующие изображениям a-f (для наглядности изображения дважды прологарифмированы); m – усредненное изображение проезда.

работке периодических структур и регионов, не содержащих характерных признаков. Поэтому разработка комбинированных методов, использующих вычисление оптического потока и сопоставление характерных признаков, устойчивых к искажениям яркости, представляется наиболее перспективной.

В рамках этой концепции предлагается комбинированный метод, совмещающий изложенный выше алгоритм на основе оптического потока с оценкой движения путем сопоставления преимущественно вертикальных линий краевых границ (т.е. непрерывных линий, состоящих из локальных максимумов и минимумов яркости изображения (roof edges) и локальных максимумов и минимумов производной изображения $\partial I/\partial y$ (step edges) *(рис. 9, 10)*. Быстрая техника нахождения таких линий границ (за линейное время от площади изображения) изложена нами ранее [17]. Для оценки смещения используется схема голосования, в которой все пары точек $p_{I(t)} = (x_t, y) \in I(t)$ и $x_{I(t+1)} = x_{t+1}, y) \in I(t)$ границ соответствующих четырех типов (локальные минимумы и максимумы вертикальной производной, локальные минимумы и максимумы яркости) пары последовательных кадров $I(t)$, $I(t + 1)$, имеющих одинаковую вертикальную координату y, голосуют за смещение $u(x_{I(t)}, x_{I(t+1)}) = x_{t+1} - x_t$.

Эта схема голосования за параметр смещения u аналогично изложенной выше схеме на основе оптического потока порождает гистограмму распределения оценок смещения u_{fg}, на которой присутствуют два выраженных локальных экстремума, соответствующих нулевому полю движения фона и ненулевому полю движения объекта. Визуализация работы алгоритма определения смещения на основе преимущественно вертикальных границ представлена на *рисунке 10*.

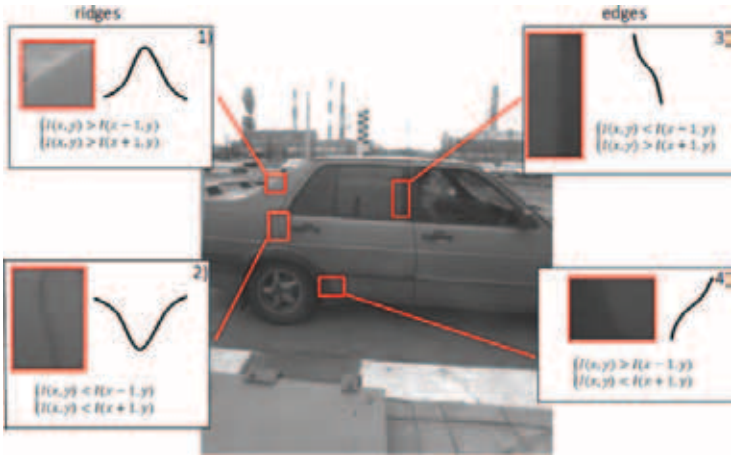


Рис. 9. Четыре типа преимущественно вертикальных границ.

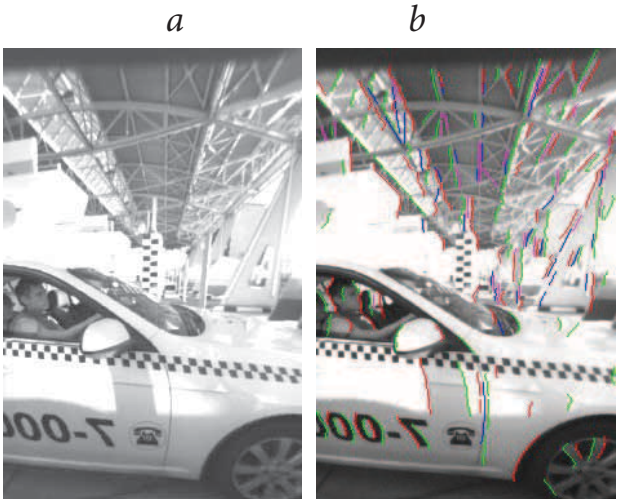


Рис. 10. *Визуализация работы алгоритма нахождения преимущественно вертикальных границ на примере кадра с изображением ТС: a – исходное изображение, b – результаты детектирования границ (малиновым цветом обозначены границы типа 1, синим – типа 2, красным – типа 3, зеленым – типа 4).*

Пусть $I(t)$, $I(t + 1)$ – пара последовательных кадров видеоряда. Тогда предлагаемый алгоритм оценки параметров движения объекта в рамках кулисной модели сцены в момент времени t представляется следующим образом:

1. Оценка оптического потока.

1.1. Изображения $I(t)$, $I(t + 1)$ сглаживаются с ядром Гаусса.

1.2. В каждой точке изображения $I(t)$, для которой выполняется условие (7) $p \in Q$, определяется оптический поток по формуле (6).

1.3. Оптический поток, вычисленный в каждой точке p проверяется на допустимость в условиях ограничений на скорость, задаваемых априорно $u_{min} \leq u(p) \leq u_{max}$.

1.4. Строится гистограмма $h(u)$ (с заданным шагом дискретизации – параметр алгоритма) голосований всех точек за величину смещения u.

1.5. Гистограмма $h(u)$ сглаживается с ядром Гаусса для получения устойчивых низкочастотных экстремумов.

1.6. Проверка достоверности полученных оценок параметров движения.

1.7. Если гистограмма содержит выраженный локальный экстремум в точке $x = 0$, т.е. $h(0) \geq bg_{min}$, где bg_{min} – параметр метода, оценка движения считается достоверной, goto 2 (переход к выполнению шага 2 алгоритма). Иначе делается вывод о наличии существенных искажений яркости на текущей последовательности изображений, так как оптический поток фона не был надежно оценен, goto 4 (переход к выполнению шага 4 алгоритма).

2. $u_{nz} = \text{argmax}_{h(u)} u, u \neq 0, u \in [u_{min}, u_{max}]$, где u_{min}, u_{max} – параметры метода, минимальное и максимальное мгновенное смещение объекта соответственно.

3. Проверка гипотезы о том, что u_{nz} соответствует смещению движущегося объекта в момент времени t. Если $|u_{nz}| \geq fg_{min}$, где fg_{min} – параметр алгоритма, минимальное число точек, проголосовавших за смещение u_{nz}, достаточное, чтобы оно принималось в качестве параметра поля движения объекта, то $u_{fg} = u_{nz}$, иначе принимается гипотеза о том, что $u_{fg} = u_{bg} = 0$, то есть движение в момент времени t в наблюдаемой сцене отсутствует. Выход.

4. Оценка параметров движения на основе преимущественно вертикальных границ.

4.1. На паре изображений $I(t)$, $I(t + 1)$ детектируются преимущественно вертикальные линии четырех типов $E(I(t))_{(1)-(4)}$, $E(I(t + 1))_{(1)-(4)}$ *(рис. 9).*

4.2. Построение гистограммы голосования всех значащих точек за параметры движения. Каждая пара точек $xg_k^t \in E(I(t))_k = (x^t, y^t)$, $xg_k^{t+1} = (x^{t+1}, y^{t+1}) \in E(I(t+1))_k$, $k \in$ (1)-(4) таких, что $y^t = y^{t+1}$, голосует за смещение $u' = x^{t+1} - x^t$ при условии, что $u_{min} \leq u' \leq u^{max}$; goto 2.

В качестве начальной инициализации алгоритма прослеживания объектов используются срабатывания обучаемого детектора объектов. На *рисунке 11* продемонстрированы результаты работы алгоритмов оценки параметров движения на основе оптического потока и вертикальных границ в условиях нестабильного освещения.

DOI: 10.22204/2410-4639-2016-092-04-17-37 № 4 (92) октябрь–декабрь 2016 г.

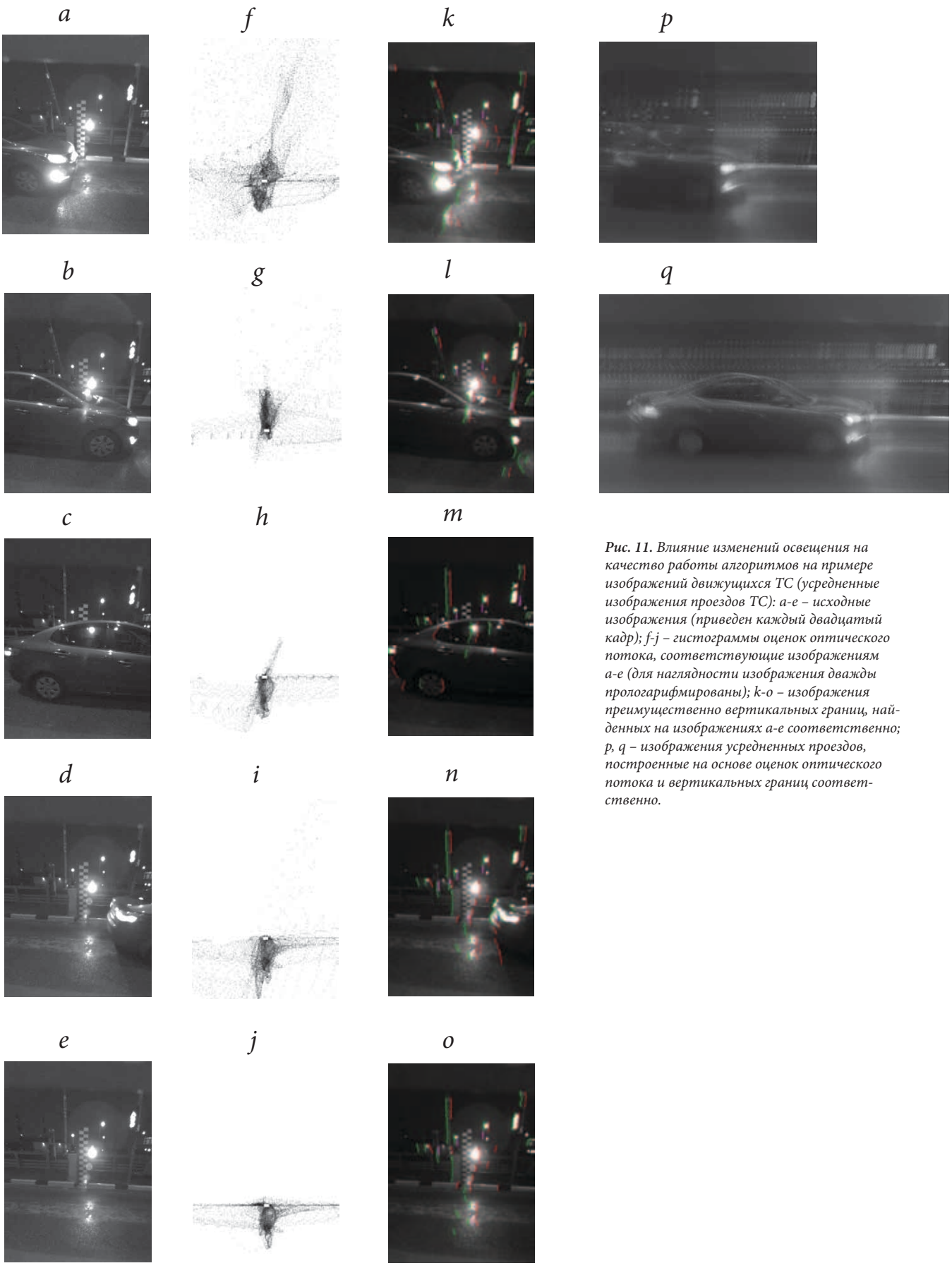


Рис. 11. *Влияние изменений освещения на качество работы алгоритмов на примере изображений движущихся ТС (усредненные изображения проездов ТС): a-e – исходные изображения (приведен каждый двадцатый кадр); f-j – гистограммы оценок оптического потока, соответствующие изображениям a-e (для наглядности изображения дважды прологарифмированы); k-o – изображения преимущественно вертикальных границ, найденных на изображениях a-e соответственно; p, q – изображения усредненных проездов, построенные на основе оценок оптического потока и вертикальных границ соответственно.*

Отметим, что все использованные нами методы имеют линейную сложность от размеров изображения, поэтому предложенный алгоритм также имеет линейную сложность. Это позволяет применять его в режиме реального времени для прослеживания движущихся объектов на каждом кадре видеоряда.

Для оценки производительности алгоритма в АКТС был реализован модуль определения направления въезда и выезда ТС в поле зрения камеры на основе предложенного метода. На тестовом наборе данных (~260 часов видео (40 кадров в секунду), полученных в режиме онлайн, содержащих ~25000 проездов ТС) качество определения направления ТС составило ~99.991%. Данный модуль в настоящее время используется в промышленной версии АКТС, работающей в реальных условиях, где при любых погодных условиях и освещенности наблюдаются схожие средние показатели качества определения направления, что демонстрирует высокую надежность предложенного алгоритма.

Заключение

Представлена структурная модель модуля инкрементного неконтролируемого обучения классификатора для распознающих систем реального времени на основе статического классификатора и модели событий высокого уровня, обеспечивающего робастность к ошибкам классификатора и модели при помощи построения проверки согласованности меток для кластеров схожих объектов. Рассмотрен вопрос спецификации этой модели для задачи детектирования движущихся объектов на видео. В качестве статического детектора было предложено использовать обобщающую древовидную модификацию алгоритма Виолы–Джонса, имеющую сопоставимую с классическим алгоритмом вычислительную сложность, но, в отличие от него, обеспечивающую возможность обучения на новых данных без сохранения исходной обучающей выборки, что решает проблему хранения больших объемов данных на промышленных вычислителях. Описаны характерные для реальных систем проблемы обучения детектора: наличие перепадов яркости на распознаваемых изображениях и недостаточных данных в обучающей выборке для начального обучения. Предложены варианты их решения: модификация признаков, используемых классификатором, и аугментация данных для обучения. В качестве модели эволюции детектируемого объекта предложена кулисная модель движения. Для ее реализации разработан комбинированный алгоритм на основе оценок оптического потока и сопоставления преимущественно вертикальных границ изображения, имеющий низкую вычислительную сложность и высокую робастность к изменениям яркости. Приведены вычислительные эксперименты, демонстрирующие надежность предложенных методов и техник. Каждый из них может быть использован вне контекста инкрементного обучения для надежного решения задач детектирования образов и прослеживания движущихся объектов в видеопотоке. Все описанные алгоритмы в настоящее время реализованы и применяются в промышленной версии разработанной коллективом статьи системы автоматической классификации транспортных средств, демонстрирующей в среднем высокое качество (>99.7%) и надежность работы в любых штатных условиях.

DOI: 10.22204/2410-4639-2016-092-04-17-37 № 4 (92) октябрь–декабрь 2016 г.

Литература

1. ***А.А. Иванова, Е.Г. Кузнецова, Д.П. Николаев***
В *Сб. труд. 39-й Междисципл. шк.-конф. ИТиС 2015 «Информационные технологии и системы 2015», (РФ, Сочи, 7–11 сентября, 2015 г.),* Москва, Изд. ИППИ им. А.А. Харкевича РАН, 2015, с. 1169–1184.
2. ***G.M. Weiss, F. Provost***
JAIR, 2003, **19**, 315. DOI: 10.1613/jair.1199.
3. ***X.Y. Liu, J. Wu, Z.H. Zhou***
IEEE Transactions on Systems, Man, and Cybernetics — Part B: Cybernetics, 2009, **39**(2), 539. DOI: 10.1109/TSMCB.2008.2007853.
4. ***T. Jo, N. Japkowicz***
ACM SIGKDD Explorations Newsletter: Special Issue on Learning from Imbalanced Datasets, 2004, **6**(1), 40. DOI: 10.1145/1007730.1007737.
5. ***G.E.A.P.A. Batista, R.C. Prati, M.C. Monard***
ACM SIGKDD Explorations Newsletter: Special Issue on Learning from Imbalanced Datasets, 2004, **6**(1), 20. DOI: 10.1145/1007730.1007735.
6. ***Z.H. Zhou, X.Y. Liu***
IEEE Transactions on Knowledge & Data Engineering, 2006, **18**(1), 63. DOI: 10.1109/TKDE.2006.17.
7. ***Z. Kalal, K. Mikolajczyk, J. Matas***
IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, **34**(7), 1409. DOI: 10.1109/TPAMI.2011.239.
8. ***O. Javed, S. Ali, M. Shah***
В *Proc. CVPR 2005: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2-Vol. Ed., (USA, CA, San-Diego, 20–25 June, 2005),* Vol. 1, IEEE Publ., 2005, pp. 696–701. DOI: 10.1109/CVPR.2005.259.
9. ***D. Fortun, P. Bouthemy, C. Kervrann***
Computer Vision and Image Understanding, 2015, **134**, 1. DOI: 10.1016/j.cviu.2015.02.008.
10. ***M. Muja, D.G. Lowe***
В *Proc. 4th International Conference on Computer Vision Theory and Applications: VISAPP, (VISIGRAPP 2009), 2-Vol. Ed., (Portugal, Lisboa, 5–8 February, 2009),* Vol. 1, Portugal, Setúbal, SCITEPRESS Publ., 2009, pp. 331–340. DOI: 10.5220/0001787803310340.
11. ***A. Minkina, D. Nikolaev, S. Usilin, V. Kozyrev***
В *Proc. SPIE 9445: Seventh International Conference on Machine Vision (ICMV 2014), (Italy, Milan, 19–21 November, 2014),* SPIE Publ., 2015, 944517. DOI: 10.1117/12.2180941.
12. ***Y. Freund, R.E. Schapire***
В *Computational Learning Theory: Proc. 2nd European Conference (EuroCOLT'95), (Spain, Barcelona, 13–15, March, 1995), Ser. «Lecture Notes in Computer Science»,* **904**, Springer-Verlag Publ., 1995, pp. 23–37. DOI: 10.1007/3-540-59119-2_166.
13. ***C.P. Papageorgiou, M. Oren, T. Poggio***
В *Proc. ICCV'98: 6th International Conference on Computer Vision, (India, Bombay, 4–7 January, 1998),* IEEE Computer Society – Narosa Publ. House, 1998, pp. 555–562. DOI: 10.1109/ICCV.1998.710772.
14. ***T. Khanipov, I. Koptelov, A. Grigoryev, E. Kuznetsova, D. Nikolaev***
В *Proc. SPIE 9445: Seventh International Conference on Machine Vision (ICMV 2014), (Italy, Milan, 19–21 November 2014),* SPIE Publ., 2015, 944511. DOI: 10.1117/12.2181557.
15. ***I. Konovalenko, E. Kuznetsova***
В *Proc. SPIE 9445: Seventh International Conference on Machine Vision (ICMV 2014), (Italy, Milan, 19–21 November, 2014),* SPIE Publ., 2015, 94451C. DOI: 10.1117/12.2181226.
16. ***I.A. Konovalenko, A.B. Miller, B.M. Miller, D.P. Nikolaev***
В *Proc. 29th European Conference on Modelling and Simulation (ECMS 2015), (Bulgaria, Albena (Varna), 26–29 May, 2015),* Germany, Sbr.-Dudweiler, Digitaldruck Pirrot GmbH Publ., 2015, pp. 489–505. DOI: 10.7148/2015-0499.
17. ***А.А. Белогаев, Е.Г. Кузнецова, Д.П. Николаев***
В *Сб. труд. 38-й Конф.-шк. ИППИ РАН ИТиС 2014 «Информационные технологии и системы 2014», (РФ, Нижний Новгород, 1–5 сентября, 2014 г.),* Москва, Изд. ИППИ им. А.А. Харкевича РАН, 2014, с. 184–189.

English

Development of Computer Vision Algorithms of Incremental Unsupervised Learning for Detection of Complex Structured Moving Objects*

Elena G. Kuznetsova –
A.A. Kharkevich Institute
for Information Transmission Problems RAS
19-1, Bolshoy Karetny Per., Moscow, 127051, Russia
e-mail: vojageur@gmail.com

Igor V. Polyakov –
A.A. Kharkevich Institute
for Information Transmission
Problems RAS
19-1, Bolshoy Karetny Per.,
Moscow, 127051, Russia
e-mail: polyakov@visillect.com

Dmitriy P. Nikolaev –
A.A. Kharkevich Institute
for Information Transmission
Problems RAS
19-1, Bolshoy Karetny Per.,
Moscow, 127051, Russia
e-mail: d.p.nikolaev@gmail.com

Dmitriy N. Matsnev –
A.A. Kharkevich Institute
for Information Transmission
Problems RAS
19-1, Bolshoy Karetny Per.,
Moscow, 127051, Russia
e-mail: matsnev@iitp.ru

* *The work was financially supported by RFBR (project 13-01-12106).*

Abstract

The paper considers the incremental learning approach to solving the representativeness problem of training data sets for machine learning. Authors described the structural model of unsupervised incremental learning process of the pattern recognition machine while *a priory* model of the observed scene evolution was used as a teacher. An option of extra-learning model for the task of moving objects detection in the videostream along with a set of necessary computer vision algorithms is suggested. The tree-form modification of the Viola–Jones algorithm has been used as the trained detector of moving objects, and a number of techniques were proposed to enhance the model's learning and extra-learning quality. The model of translational motion was applied as the evolution model. To implement this model the new tracking images method, based on the combination of the optical flow estimation and the comparison of orthotropic borders on images, was proposed. The results of computational experiments demonstrated the reliability of the proposed approaches.

Keywords: machine learning, unsupervised learning, incremental learning, Viola–Jones, object tracking.

Images & Tables

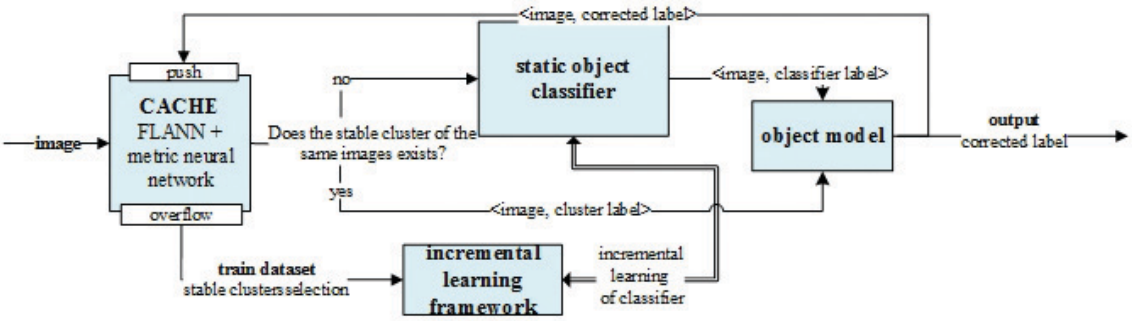


Fig. 1. Unsupervised learning framework.

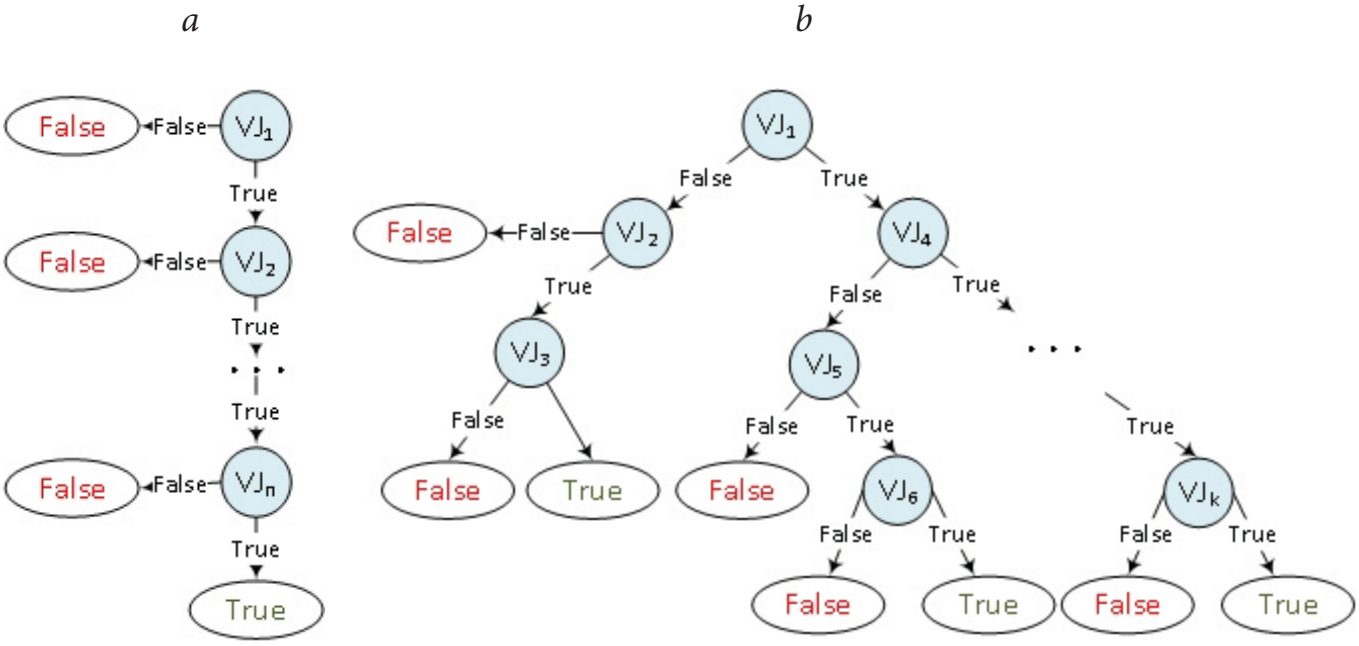


Fig. 2. Viola-Jones structure: a – classical cascade, b – tree-like classifier.

DOI: 10.22204/2410-4639-2016-092-04-17-37 № 4 (92) октябрь–декабрь 2016 г.

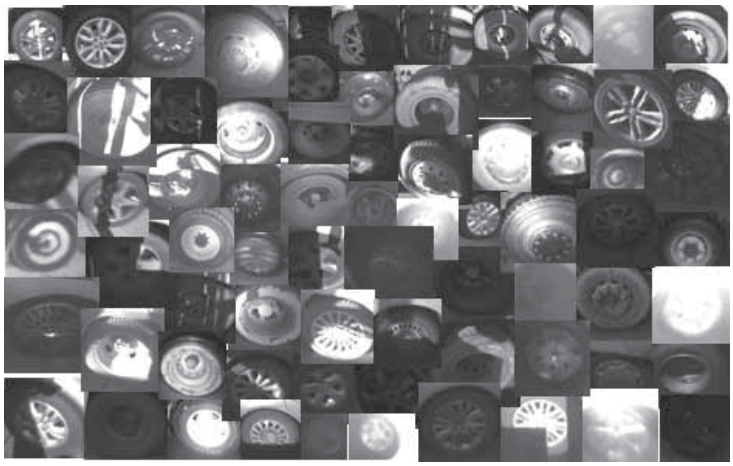


Fig. 3. Examples of a test dataset.

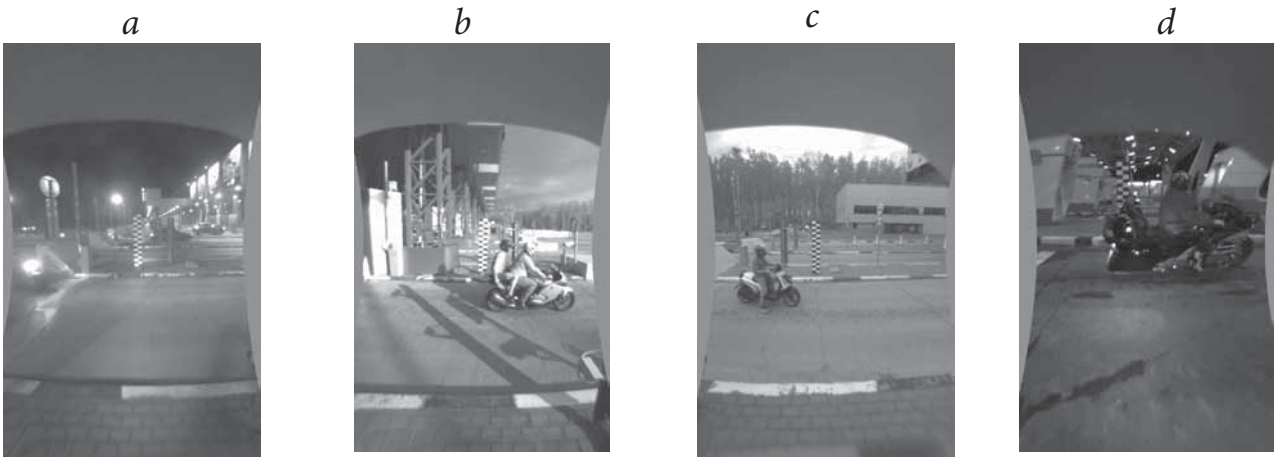


Fig. 4. Examples of AVC frames containing images of motorcycle (a-d).

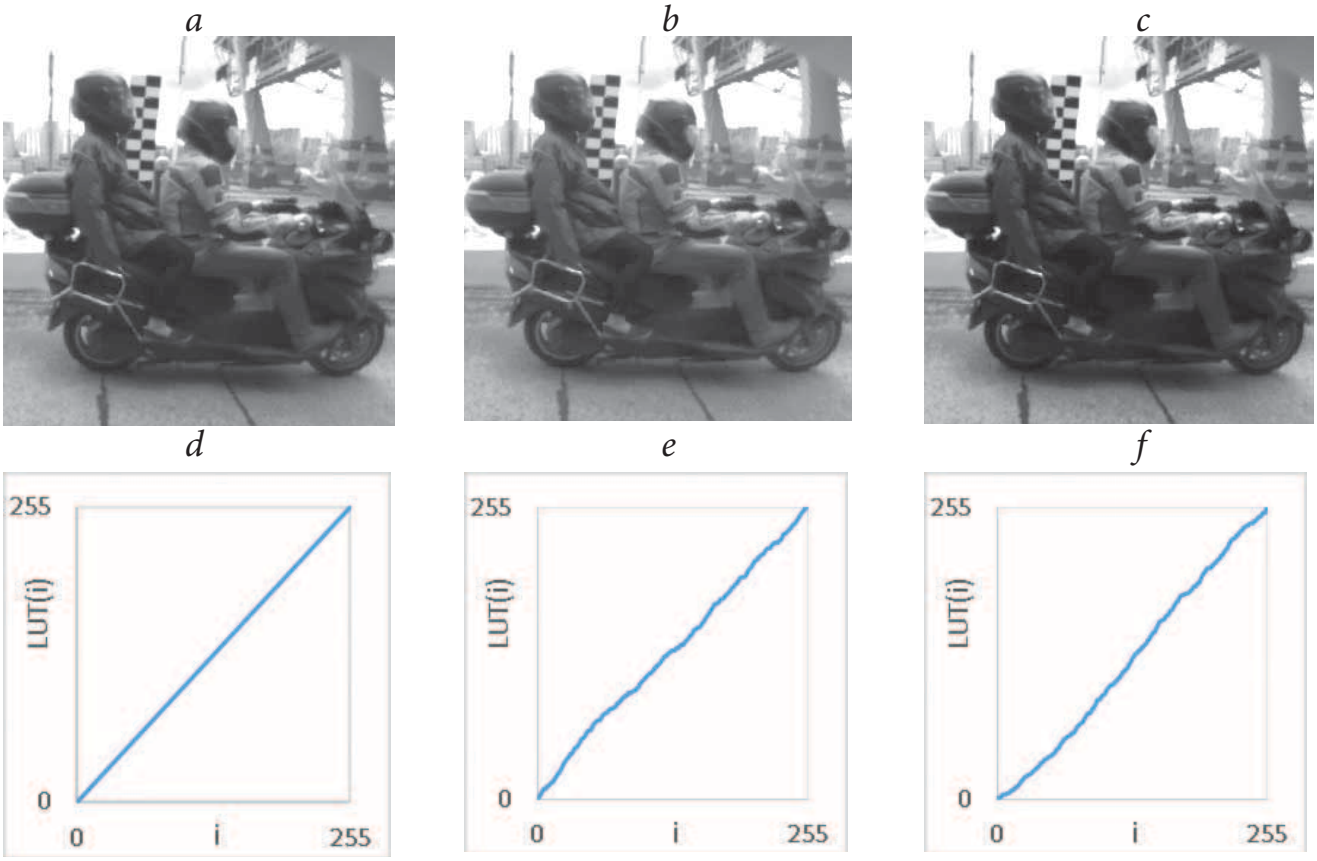


Fig. 5. Examples of images augmentation through modification of brightness scale: a – source image, b,c – modified images, d-f – brightness scales.

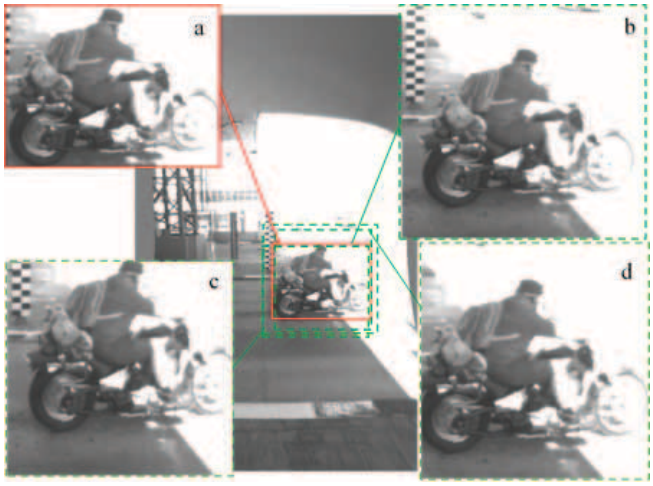


Fig. 6. Examples of images augmentation through offset the object's borders: a – the original region of the object, b-d – examples of modification.

Table 1. *Results of the motorcycle detectors training*

Training using raw data (without augmentation)				Training using data synthesis (with augmentation)			
Level of the tree-like structure	Type 2 error	Type 1 error	N_{wc}	Level of the tree-like structure	Type 2 error	Type 1 error	N_{wc}
1	0.0210	0.3721320	6	1	0.0110	0.406106	7
2	0.0355	0.2280660	9	2	0.0115	0.343031	7
3	0.0590	0.1193880	15	3	0.0225	0.18222	19
4	0.0645	0.1002590	10	4	0.0360	0.118939	20
5	0.0725	0.0958354	12	5	0.0430	0.0716792	24
6	0.086	0.0909151	19	6	0.0610	0.048532	19
7	0.0965	0.0754331	27	7	0.0715	0.0309049	24
8	0.1000	0.0798331	16	8	0.0792	0.0204564	21
9	0.1335	0.0408132	20	9	0.0860	0.0114894	21
10	0.140	0.0360354	23	-	-	-	-
11	0.1415	0.0339197	10	-	-	-	-
12	0.1470	0.0280738	8	-	-	-	-
13	0.1510	0.0232482	11	-	-	-	-
14	0.1605	0.0179654	22	-	-	-	-
15	0.1605	0.0143202	4	-	-	-	-
16	0.1645	0.0136899	21	-	-	-	-

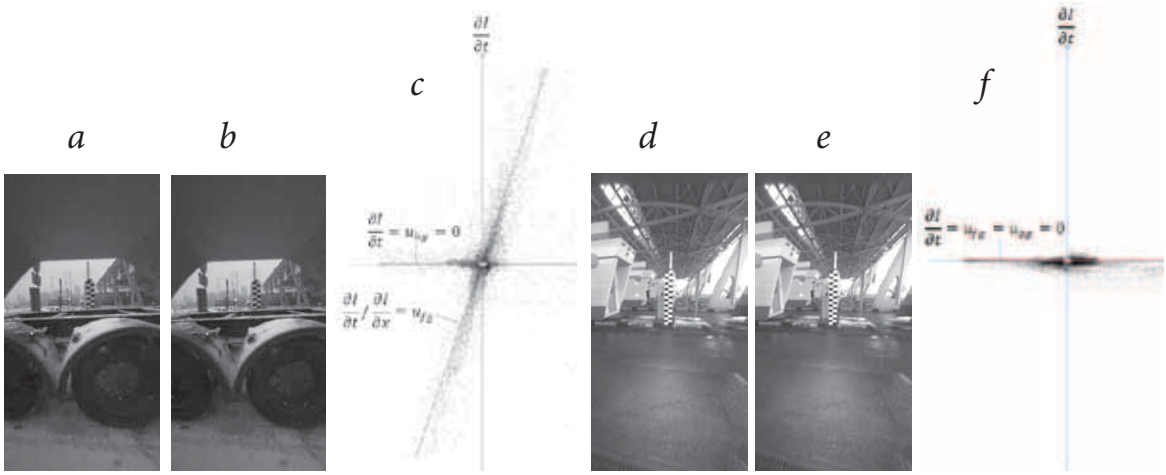


Fig. 7. Examples of estimation histograms of the optical flow ∂I/∂t(∂I/∂x) for image I at time: a – moving object, b – static scene (a, d – I(t), b, e – I(t+1), c, f – ∂I/∂t(∂I/∂x)).

Fig. 8. Visualization of the algorithm based on optical flow estimates for determining the field parameters of the moving vehicle in the video sequence: a-f – initial video-images, g-l – optical flow estimation histograms, m – average image of the vehicle passing.

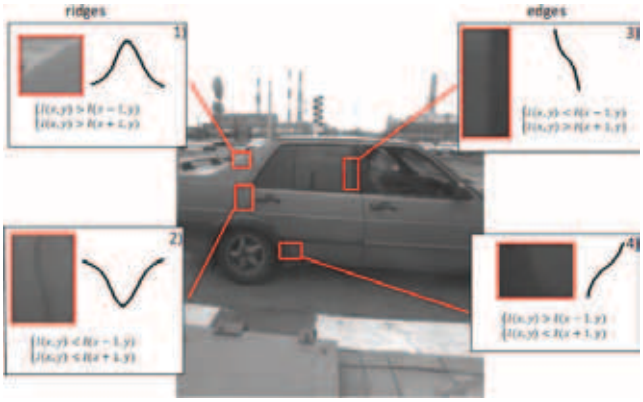


Fig. 9. The four types of predominantly vertical boundaries.

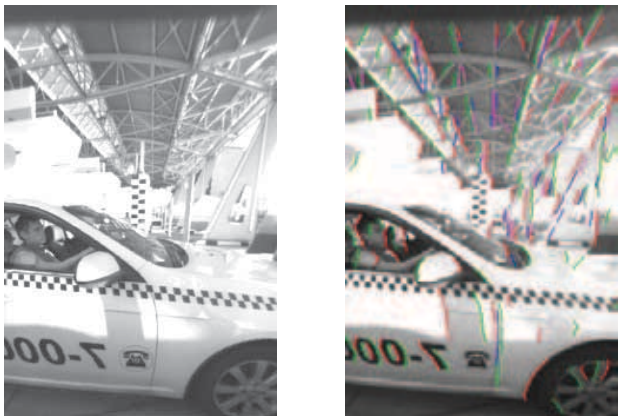


Fig. 10. Detection of predominantly vertical boundaries (crimson, blue, red and green colours correspond to boundaries of type 1), 2), 3) and 4) respectively).

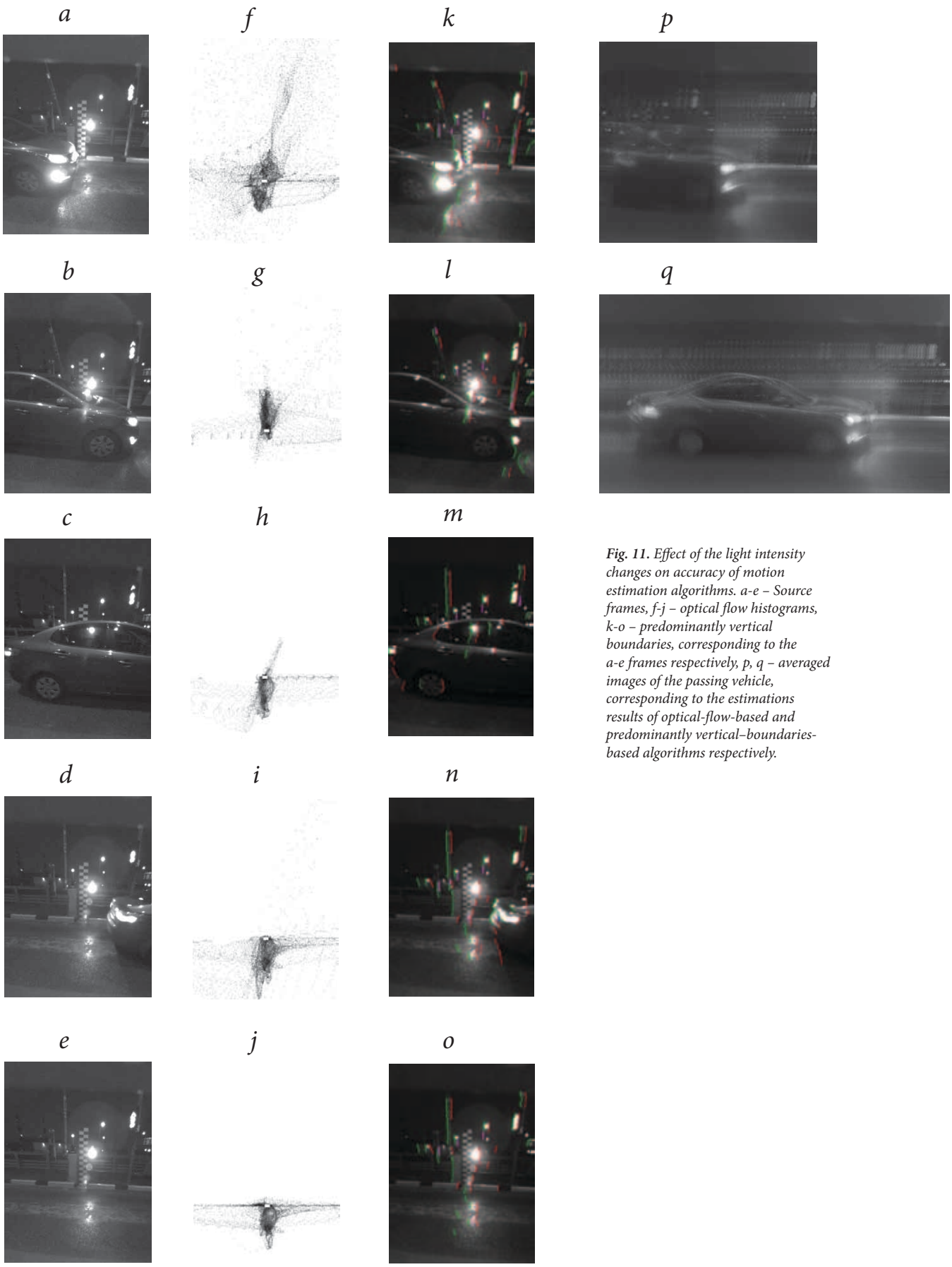


Fig. 11. *Effect of the light intensity changes on accuracy of motion estimation algorithms. a-e – Source frames, f-j – optical flow histograms, k-o – predominantly vertical boundaries, corresponding to the a-e frames respectively, p, q – averaged images of the passing vehicle, corresponding to the estimations results of optical-flow-based and predominantly vertical–boundaries-based algorithms respectively.*

DOI: 10.22204/2410-4639-2016-092-04-17-37 № 4 (92) октябрь–декабрь 2016 г.

References

1. ***1. A.A. Ivanova, E.G. Kuznetsova, D.P. Nikolaev***
In *Proc. 39th School-Conf. ITᵃS 2015 "Information Technologies and Systems 2015" [Informatsionnye tekhnologii i sistemy], (RF, Sochi, 7–11 September, 2015),* RF, Moscow, A.A. Kharkevich IITP RAS Publ., 2015, pp. 1169-1184 (in Russian).
2. ***G.M. Weiss, F. Provost***
JAIR, 2003, **19**, 315. DOI: 10.1613/jair.1199.
3. ***X.Y. Liu, J. Wu, Z.H. Zhou***
IEEE Transactions on Systems, Man, and Cybernetics — Part B: Cybernetics, 2009, **39**(2), 539. DOI: 10.1109/TSMCB.2008.2007853.
4. ***T. Jo, N. Japkowicz***
ACM SIGKDD Explorations Newsletter: Special Issue on Learning from Imbalanced Datasets, 2004, **6**(1), 40. DOI: 10.1145/1007730.1007737.
5. ***G.E.A.P.A. Batista, R.C. Prati, M.C. Monard***
ACM SIGKDD Explorations Newsletter: Special Issue on Learning from Imbalanced Datasets, 2004, **6**(1), 20. DOI: 10.1145/1007730.1007735.
6. ***Z.H. Zhou, X.Y. Liu***
IEEE Transactions on Knowledge & Data Engineering, 2006, **18**(1), 63. DOI: 10.1109/TKDE.2006.17.
7. ***Z. Kalal, K. Mikolajczyk, J. Matas***
IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, **34**(7), 1409. DOI: 10.1109/TPAMI.2011.239.
8. ***O. Javed, S. Ali, M. Shah***
In *Proc. CVPR 2005: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2-Vol. Ed., (USA, CA, San-Diego, 20–25 June, 2005),* Vol. 1, IEEE Publ., 2005, pp. 696–701. DOI: 10.1109/CVPR.2005.259.
9. ***D. Fortun, P. Bouthemy, C. Kervrann***
Computer Vision and Image Understanding, 2015, **134**, 1. DOI: 10.1016/j.cviu.2015.02.008.
10. ***M. Muja, D.G. Lowe***
In *Proc. 4th International Conference on Computer Vision Theory and Applications: VISAPP, (VISIGRAPP 2009), 2-Vol. Ed., (Portugal, Lisboa, 5–8 February, 2009),* Vol. 1, Portugal, Setúbal, SCITEPRESS Publ., 2009, pp. 331–340. DOI: 10.5220/0001787803310340.
11. ***A. Minkina, D. Nikolaev, S. Usilin, V. Kozyrev***
In *Proc. SPIE 9445: Seventh International Conference on Machine Vision (ICMV 2014), (Italy, Milan, 19–21 November, 2014),* SPIE Publ., 2015, 944517. DOI: 10.1117/12.2180941.
12. ***Y. Freund, R.E. Schapire***
In *Computational Learning Theory: Proc. 2nd European Conference (EuroCOLT'95), (Spain, Barcelona, 13–15, March, 1995), Ser. «Lecture Notes in Computer Science»,* **904**, Springer-Verlag Publ., 1995, pp. 23–37. DOI: 10.1007/3-540-59119-2_166.
13. ***C.P. Papageorgiou, M. Oren, T. Poggio***
In *Proc. ICCV'98: 6th International Conference on Computer Vision, (India, Bombay, 4–7 January, 1998),* IEEE Computer Society – Narosa Publ. House, 1998, pp. 555–562. DOI: 10.1109/ICCV.1998.710772.
14. ***T. Khanipov, I. Koptelov, A. Grigoryev, E. Kuznetsova, D. Nikolaev***
In *Proc. SPIE 9445: Seventh International Conference on Machine Vision (ICMV 2014), (Italy, Milan, 19–21 November 2014),* SPIE Publ., 2015, 944511. DOI: 10.1117/12.2181557.
15. ***I. Konovalenko, E. Kuznetsova***
In *Proc. SPIE 9445: Seventh International Conference on Machine Vision (ICMV 2014), (Italy, Milan, 19–21 November, 2014),* SPIE Publ., 2015, 94451C. DOI: 10.1117/12.2181226.
16. ***I.A. Konovalenko, A.B. Miller, B.M. Miller, D.P. Nikolaev***
In *Proc. 29th European Conference on Modelling and Simulation (ECMS 2015), (Bulgaria, Albena (Varna), 26–29 May, 2015),* Germany, Sbr.-Dudweiler, Digitaldruck Pirrot GmbH Publ., 2015, pp. 489–505. DOI: 10.7148/2015-0499.
17. ***A.A. Belogaev, E.G. Kuznetsova, D.P. Nikolaev***
In *Proc. 38th Conf.-School ITᵃS 2014 "Information Technologies and Systems 2014" [Informatsionnye tekhnologii i sistemy], (RF, Nizhniy Novgorod, 1–5 September, 2014),* RF, Moscow, A.A. Kharkevich IITP RAS Publ., 2014, pp. 184–189 (in Russian).

Проективно инвариантное описание овалов с симметриями трех родов*

П.П. Николаев

Рассмотрены концептуальный подход к задаче проективно инвариантного распознавания плоских кривых и схема его численной реализации, позволяющие получить проективно инвариантное описание плоских овальных фигур с элементами скрытой симметрии трех родов – вращательной, радиальной (центральной) и аксиальной (осевой) – без использования с этой целью (в качестве элементов базиса) позиций каких-либо точек контура с «особыми» проективно устойчивыми характеристиками. Описание предлагается формировать, опираясь на интегральные свойства кривой, посредством вспомогательной инвариантной структуры, названной тангенциальным образом овала, вычисление которой возможно на этапе обработки, формирующей позиционную оценку «образа центра симметрии». При этом численные схемы поиска координат образа центра ранее предложены, промоделированы и описаны нами. Обсуждены методы применения вводимых инструментов анализа для инвариантного описания овалов, не обладающих элементами симметрии, при наличии дополнительной точки, фиксированной вне контура фигуры.

Ключевые слова: касательная, ортоформа, индекс ротации, плюккеровы полюс и поляра, вурф-отображение, тангенциальный образ.

** Работа выполнена при финансовой поддержке РФФИ (проекты №№ 13-01-12107 и 16-07-00836).*

Цели работы и инструменты исследования

Цель предлагаемой работы – дать конкретизацию тезиса (детализировать концепцию и обсудить схему численной ее реализации, проект РФФИ № 13-01-12107 «Методы и процедуры распознавания гладких выпуклых кривых в произвольном ракурсе»). Данные, включенные в статью, не дублируют более ранние авторские публикации этого направления, выводя заявляемую концепцию (в качестве содержательной, теоретически и практически достижимой методами дискретной обработки кривой) на уровень описания идей и процедурных шагов получения целевого «продукта» – проективно инвариантного описания гладкой выпуклой кривой (овала в произвольном ракурсе оптической регистрации, задаваемом в виде плоской центральной проекции), использующего безотносительно к типу неявной ее симметрии исключительно интегральные ее характеристики (т.е. особенности формы объекта, не требующие оценки дифференциальных инвариантов кривой в точке, с их технически недостижимым уровнем точности позиционного представления). В уже опубликованных в работах нами по проекту № 13-01-12107 предложены разнообразные методы и вычислительные схемы проективно инвариантной репрезентации овалов, процедурно объединившие (в рамках двух независимых подходов) обработку кривых с радиальным и осевым типами симметрии, а для контуров со скрытой симметрией вращения сделавшие единообразным метод инвариантного их представления для любого (нечетного) индекса ротации [1–3]. Цель настоящей работы – показать осуществимость единого процедурного подхода, дающего возможность распространить использование вспомогательной структуры тангенциального (дуального) образа кривой, предложенной нами для описания кривых ротационного типа, в качестве подходящего инструмента презентации овалов с аксиальной и радиальной симметрией.

Введем определения употребляемых далее терминов и сокращенные обозначения в списке типов симметрии и наборе инструментов анализа кривой. Типам симметрии дадим обозначения: ротационному – R, центральному – C, аксиальному – A. Ортоформой будем называть любую подходящую проекцию овала, для которой признаки ее симмет-

НИКОЛАЕВ
Петр Петрович
Институт проблем передачи информации
им. А.А. Харкевича РАН

ричной организации представлены в явном виде (они удовлетворяют правилам, формулируемым для декартовой системы координат), т.е. хорды *A*-овала, ортогональные оси симметрии, рассекаются ею пополам; а хорды *C*-овала рассекаются пополам точкой центра симметрии. Ортоформой R_N-овала с индексом ротации *N* назовем кривую, которая не меняется при повороте вокруг центра ротации любого ее фрагмента угловой длины $L(2\pi/N)$ на угол, кратный $2\pi/N$. При четном *N* ортоформе присущи свойства *R* и *C* (а если существует такое разбиение на фрагменты, где фрагмент осесимметричен, то ортоформа будет иметь и *N* осей симметрии). Упорядоченный набор *N* точек R_N-овала, разрезающих его на *N* фрагментов ротации, назовем ансамблем корреспонденции (АК), откуда следует, что для любой точки P_1 его контура найдутся (*N*–1) точек с теми же проективно инвариантными (дифференциальными) свойствами, комплектующие АК. Под воздействием проективного преобразования плоскости овала в декартовом 3D-пространстве (что является моделью смены ракурса сенсорной регистрации при постановке технической задачи распознавания) явные свойства симметрии, присущие его ортоформе, могут оказаться утраченными, и для поиска элементов симметрии (ЭС) овала понадобится привлечение проективных аналогов из списка признаков симметрии (*R*-, *C*- и *A*-типов кривой), каковые нами были определены и использованы в численных моделях поиска ЭС. Этот инструментарий включает оценку вурфа (двойного отношения для коллинеарного квартета точек [4]) и исчерпывается анализом поведения касательных в рамках плюккерова полюс-полярного соответствия прямой и точки (как поляры и полюса), предложенного Плюккером для коник [5].

Важнейшим звеном в концепции единого подхода к задаче проективно инвариантного описания овалов с ЭС является тангенциальный (дуальный) образ кривой. Данные выше дефиниции позволяют формализовать это понятие. Любому АК *R*-овала можно поставить в однозначное соответствие тангенциальный ансамбль корреспонденции, если звенья ломаной, чьи вершины образуют АК, принять за плюккеровы поляры, а набор полюсов для них считать искомым тангенциальным АК (с тем же, что и у АК, числом точек, но расположенных снаружи *R*-овала). При таком дуальном соответствии контуру *R*-овала всегда можно однозначно сопоставить охватывающую его *T*-кривую (она не обязана принадлежать семейству овалов, так как кривая для R_N-овала может иметь *N* вогнутостей, оставаясь гладко замкнутой в проективном смысле), образованную однопараметрическим семейством плюккеровых полюсов для хорды – поляры, являющейся у ортоформы стороной правильного *N*-угольника, совершающего поворот на угол $2\pi/N$ вокруг центра ротации *O* (центр *O* ортоформы – это точка явного центра, в общем же случае ротация идет вокруг образа центра *O*′, притом что полигон АК перестает быть правильным). Каждая из сторон полигона АК (меняющего размер при повороте вокруг неизменного центра) задает свою *N*-ую часть охватывающей *T*-кривой, которые сомкнутся в гладком сопряжении при завершении поворота на угол $2\pi/N$. Такую *T*-кривую и назовем *T*-образом овала, где АК случайной выборки, образованный триадой точек *P*1, *P*2, *P*3, делит *R*3-овал на сопряженные фрагменты, показанные на *рисунке 1* линиями разного цвета, что задает композицию тангенциального АК в виде триады *T*1, *T*2, *T*3, принадлежащей его *T*-образу. Слева пунктирными прямыми показан треугольник *t*1, *t*2, *t*3, для которого тангенциальный АК принадлежит точкам вогнутости *T*-образа. В рамке справа выписаны «структурные формулы» вурф-функций (как компонент проективно инвариантного базиса, получаемого обходом обоих контуров): *w*1, *w*2 и добавочная *w*3 – в качестве примеров простой, но не универсальной процедуры формирования *W*-отображения (детекция линии горизонта *HL* обеспечивает оценку *w*3), а также функция $wR = W(O, iP, R, eP)$, вычисление которой идет по схеме универсального 11-точечного шаблона, описываемого далее. В шаблоне используется кривая *T*-образа, получаемая путем ввода *T*1 в инвариантную композицию на пересечении касательных в *P*1 и *P*2, где *P*1, «скользящая по фрагменту» (черная часть контура овала) до позиции *P*2, детерминирует координаты девяти подвижных его точек (центр *O* фиксирован, показана «правая часть» шаблона, формирующая оценку *R* для *wR*, левая ей симметрична, она строит *wL*). *Рисунок 2* на примере R_3-овала, моделируемого на *рисунке 1*, демонстрирует вид вурф-

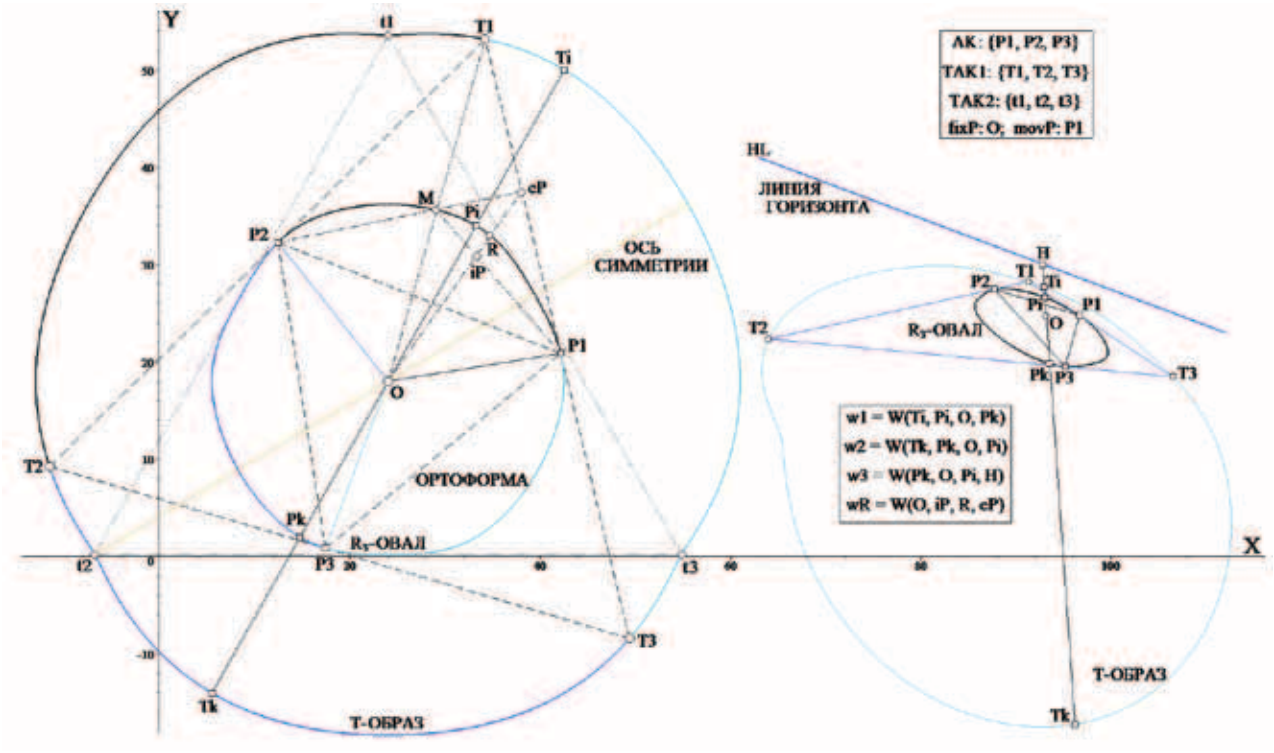


Рис. 1. Ортоформа R_3-овала и его Т-образ (слева) и они же в перспективе (справа); ТАК – тангенциальный ансамбль корреспонденции.

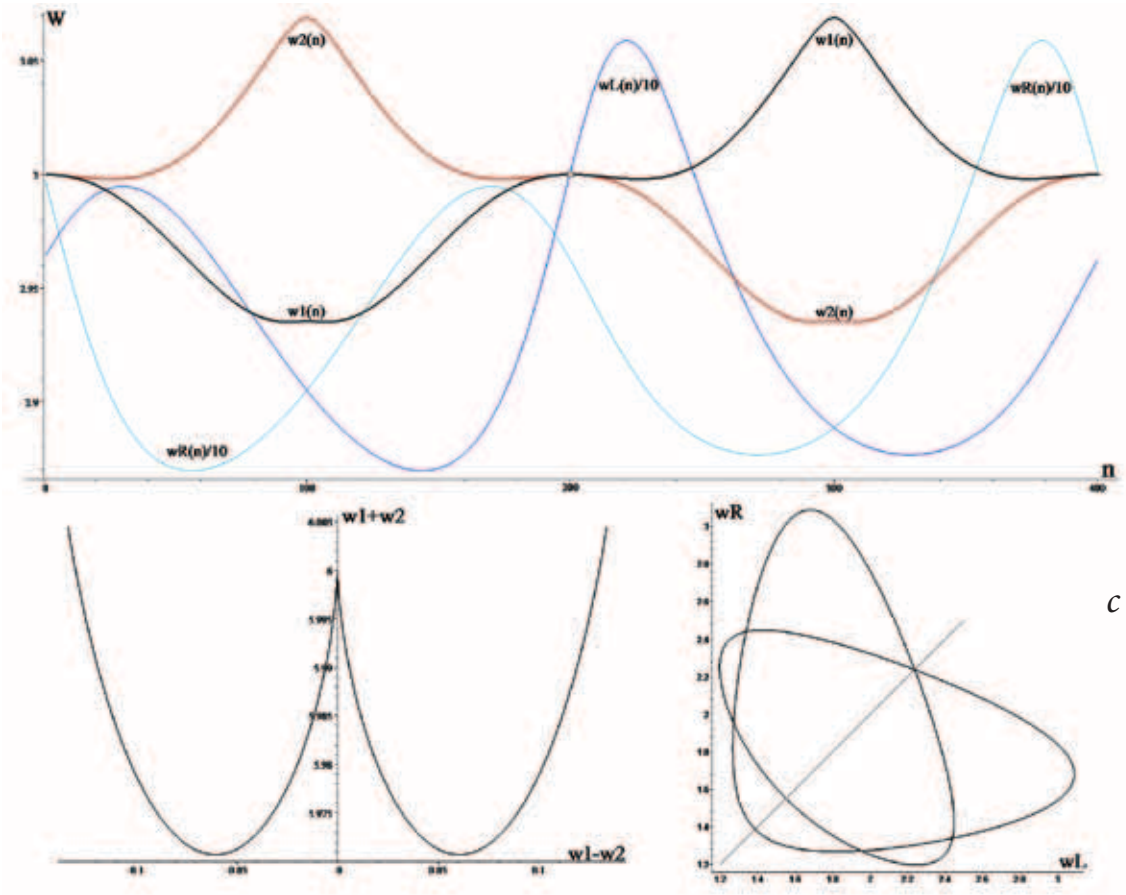


Рис. 2. Вид вурф-функций (w1, w2 и wL, wR) (a) и их W-отображений (b, c).

DOI: 10.22204/2410-4639-2016-092-04-38-54 № 4 (92) октябрь–декабрь 2016 г.

функций и *W*-отображений при «быстрой» и универсальной процедурах обработки.

Нетрудно заметить, что формально (посредством операции с поворотом фрагмента ортоформы) *C*-овал является частным случаем R_N-овала при $N = 2$. Единственное отличие – *T*-образ *C*-овала представляет собой прямую (у ортоформы – это несобственная прямая проективной плоскости овала), в общем случае ее задает линия горизонта *HL* (*рис. 1*) для некоторого конкретного ракурса оптической регистрации *C*-овала. В той же парадигме «поворота фрагмента» *A*-овал является «частным производным» *C*-овала, где один из двух фрагментов, рассекаемых осью *A*-ортоформы, дополнительно повернут на π вокруг оси в плоскости кривой. Таким образом, в цепочке смен типа симметрии $R \to C \to A$ можно усмотреть редукцию характеристик носителя *T*-образа: *T*-кривая → → прямая *HL* → уникальная точка на *HL*, а любая точка носителя во всех трех случаях является полюсом какой-либо хорды (пара касательных в концевых ее точках и задает полюс в позиции их пересечения) в качестве плюккеровой поляры. Вывод: для реализации целевой задачи нам необходимо описать процедуру, в которой однородным образом используются все точки овальной кривой с применением координат ее *T*-образа, а также позиций центров ЭС. Пока не обсуждается процедурное единообразие на этапе формирования *T*-образа: речь пойдет об информативном вкладе в проективную теорию овальных кривых с ЭС, когда для их инвариантной репрезентации используется связь распределенных характеристик кривой с ее «*T*-двойником». Для любого R_N-овала при $N > 2$ она является взаимно однозначной, т.е. в дуалитете ортоформы по кривой ее *TN*-образа можно восстановить форму R_N-овала (что похоже на дуальные отношения эволюты и эвольвенты, не являющиеся даже аффинно инвариантными, в отличие от имеющейся проективно инвариантной связи *R*-овала и его *T*-образа).

Выше (на *рис. 1*) была рассмотрена наиболее простая (в терминах вычислительной сложности) схема оценки вурф-отображения R_3-овала на плоскость (*W*1, *W*2) независимых вурфов в предположении, что помимо формы самой кривой известны ее T_3-образ и координаты центра ротации *O*. Хорда произвольной ориентации *i*, проходящая через *O*, пересекает дуальную пару кривых в четырех точках (*Ti, Pi, Pk, Tk*), что вместе с *O* дает пятиточечный коллинеарный набор, для которого любые два его квартета позволяют вычислить независимые компоненты *w*1*(i)* и *w*2*(i)*. Совершив операцией поворота хорды «сканирование» (по *i*) кривых в произвольном угловом диапазоне 2π/3 (что соответствует декартовой метрике угла для задания кривых в ортоформе), можно получить циклически замкнутое отображение R_3-овала *w*1*(w*2*)* (на *рис. 2b*), форма которого зависит от избранной композиции вурфов (из сета возможных пяти вурф-композиций). Характерной особенностью *W*-отображений (для «быстрой» и универсальной процедур) является «воспроизведение» *C* и *A* свойств симметрии, присущих ортофоме кривой. В демонстрационных целях R_3-овал, показанный на *рисунке 1*, имел дополнительно три оси симметрии (вдоль прямых *t*1-*O*, *t*2-*O* и *t*3-*O*), что и обусловило симметрию кривых *W*1(*W*2) относительно биссектрисы чисто положительного квадранта (*W*1, *W*2) (*рис. 2c*).

Для *R*-кривых с четным индексом *N* характерно проявление дополнительных свойств неявной радиальной симметрии (*C*-тип). Процедура формирования отображений *R*-овалов при $N = 4, 5,...$ ничем не отличается от случая $N = 3$, возможны лишь варианты реализации на этапе получения АК для *R*-контуров общего вида, когда детекцию АК R_3-кривых удается проводить более простыми методами, нежели для фигур при $N > 3$.

Последующие фазы вычислений (определение тангенциального АК, центра *O′* и *T*-образа) в процедурном плане неотличимы. Дадим пример инвариантной обработки *R*-овала *(рис. 3)*, не имеющего при $N = 3$ дополнительных свойств неявной осесимметричности, для чего рассмотрим новые кривые семейства R_3-овалов, разметив их как «овал 2» и «овал 3». «Оптимальная» процедура на каждом *i*-ом шаге перевычисления базиса («5-set») не привлекает координат дуальной пары кривых сверх квартета точек *T*2, *M*2, *m*2, *t*2 (*T*3, *M*3, *m*3, *t*3 для «овала 3»), тогда как по универсальной схеме («11-set») вычисляются сначала позиции *A* и B (в виде концевых точек поляры для

полюса *T*, что создает возможность позиционной оценки для *eA* и *eB*, а, стало быть, и для детекции *L* и *R*). Вариант использования функций *W(A, iA, M, eA)* и *W(B, iB, M, eB)*, «экономящий» на операциях вычисления пересечений прямых с контуром овала (имеются в виду затраты оценок *L*, *R*), не подходит по причине вырождения вурфа: согласно теории все эти функции равны константе 2/3 (*рис. 3b*, уравнения в рамке). Известен тезис, что любые пять точек на кривой образуют так называемый «плоскостной вурф», всегда сводимый к двум независимым линейным вурфам [6]. В силу этого свойства продукт базиса 11-set гарантирует формирование невырожденного *W*-отображения. Возможны равнозатратные варианты получения отображения *WL(WR)*, при одном из которых «задающей» точкой объявляется *M* (тогда для полюса *T* требуется вычислять концевые *A* и *B* его поляры), но пригодна и схема, когда в роли «задающей» выступает точка *A* (или *B*), тогда точка *T* пересечения касательных к контуру в *A* и *B* не потребует априорного знания кривой *T*-образа. Вернемся к сравнительному обзору продукции схем 5-set и 11-set.

На уровне описания вурф-функций различие продукций «5-set» и «11-set» выражается в большей их компактности для универсального базиса (по причине сильного разброса для ординат «5-set» (*рис. 4a*) для *w*2 и *w*3 пришлось вводить весовые коэффициенты 3 и 7/4; в случае «нативных» значений при едином их масштабе форму кривых для сравниваемых процедур было бы затруднительно оценить визуально). Общей чертой для *W*-отображений сравниваемых схем можно объявить присущую всем кривым квазисимметрию.

Отметим, что 2D-вурфотображения принадлежат декартовой плоскости вурфов, а потому для решения технических задач распознавания кривых семейства R_N-овалов (к примеру при классификации объектов по признаку их проективной эквивалентности) нет необходимости привлекать вычислительно громоздкие схемы метрики близости кривых при их «сортировке»: для идентификации класса овала вполне эффективно снабдить его дескриптором, включающим элементы описания, не связанные с геометрией замкнутой линии *W*-отображения. Такими элементами могут служить интегральные характеристики (наподобие площади или же числа петель и точек перегиба *W*-кривой) и позиционные параметры (координаты точек самопересечения на «*W*-эталоне» овала). Это технический аспект задачи опознания. В плане аналитических свойств *W*-кривой уместно упомянуть такую

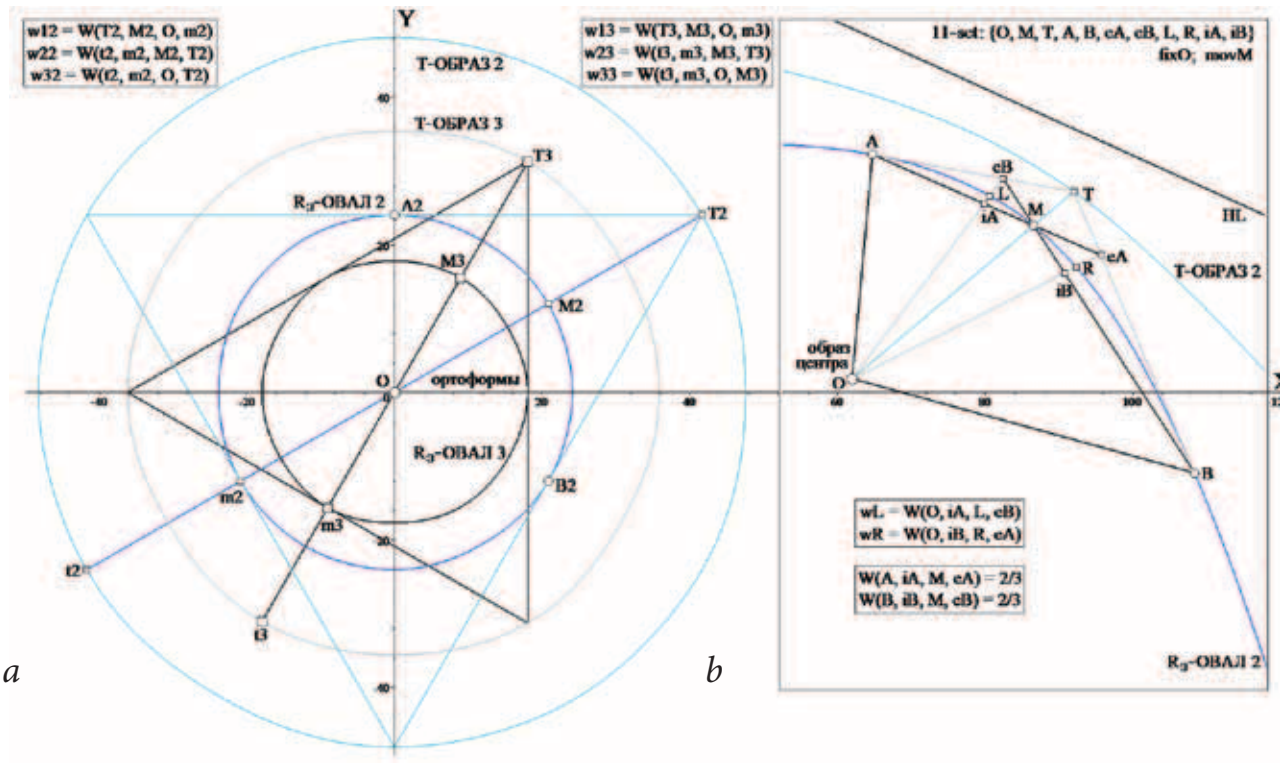


Рис. 3. R_3*-овалы («2» и «3») в ортоформе (a) и схема формирования базиса «11-set» (b).*

DOI: 10.22204/2410-4639-2016-092-04-38-54 № 4 (92) октябрь–декабрь 2016 г.

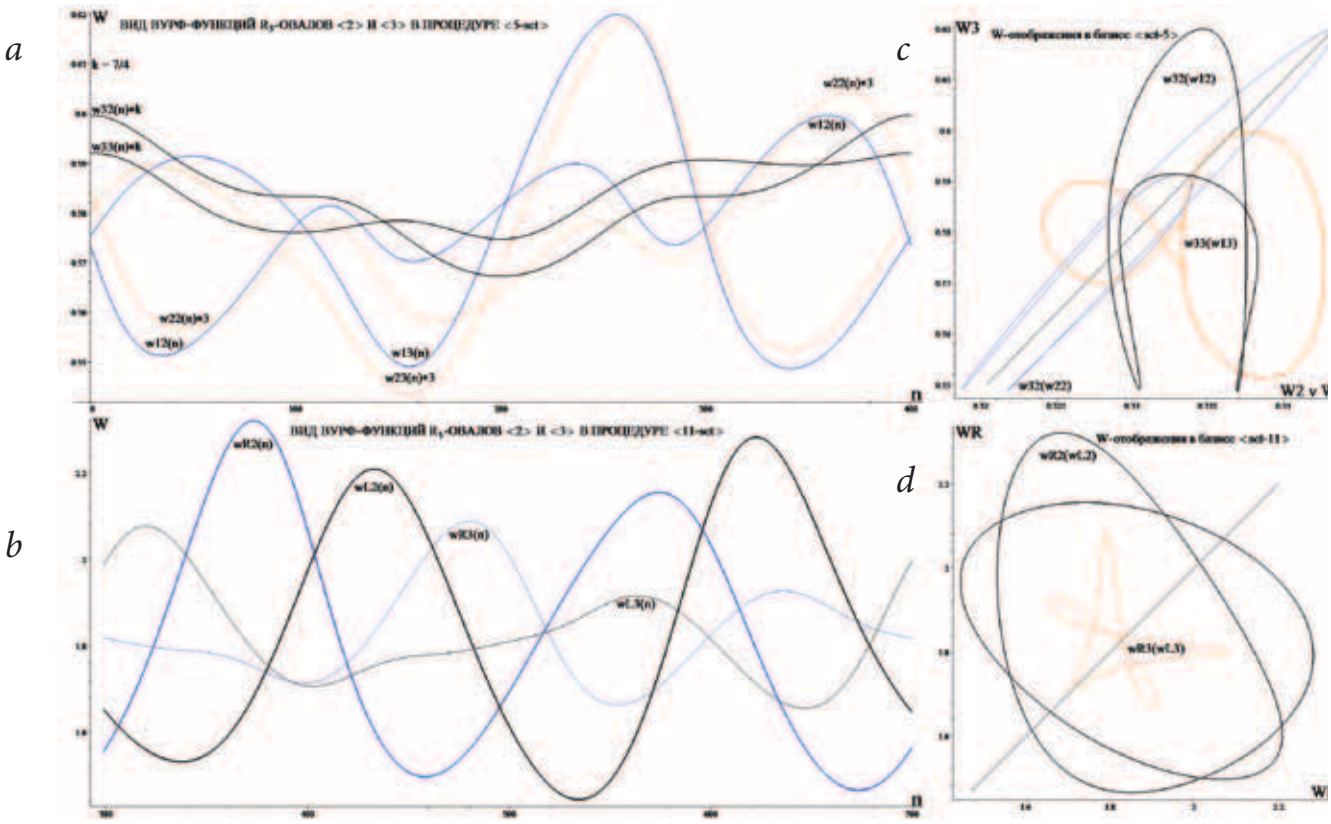


Рис. 4. *W-функции R_3-овалов в базисах «5-set», «11-set» (a,b) и вид их W-отображений (c, d).*

ее особенность: чем с большей точностью форму овала аппроксимирует семейство коник, тем меньшая площадь характерна для *W*-эталона (*W*-отображение любой коники стягивается в точку). Наконец отметим, что, принимая во внимание дуальный образ *C*-овала (прямую *HL*), схему «11-set» можно видоизменить, вычисляя *T'* (аналог *T*) в пересечении луча *O-M* не с кривой *T*-образа, а с линией горизонта *HL* (поскольку на этапе детекции *O* позиция *HL* оценивается без труда).

Метод проективно инвариантного описания *C*-овала

Радиально симметричная гладкая кривая (*C*-овал) «охватывается» линией дуального образа (прямой *HL*) таким образом, что любая хорда, пересекающая оба «объекта», уже не может снабдить блок обработки координатами пяти точек пересечения. Более того, всякая хорда, проходящая через точку образа центра, доставляет выделенный квартет таковых в качестве гармонического (т.е. модуль соответствующего вурфа всегда единичный). В силу подобного свойства *C*-овала схема формирования искомого вурф-отображения требует введения дополнительных опорных элементов, композиция которых и задаст 11-точечный шаблон «скользящего» проективного базиса (в качестве универсальной схемы, пригодной для обработки кривых любого типа симметрии). Опишем этапы его формирования на модельном примере двух *C*-овалов (*рис. 5*, черная и красная кривые), представленных ортоформой (слева) и в некотором проективном ракурсе (справа, в тех же литерных обозначениях позиций точек; разметка 11-точечного базиса «11-set» синими прямыми).

Рисунок 5 иллюстрирует важную особенность организации базиса, использующего пять точек контура овала (*A, B, M, L, R* в композиции «11-set») для описания «текущей» *M*, и «минимизированной» его версии с семью точками шаблона и тремя точками контура (*A, B, M*). Упрощенная процедура продуцирует вурф-функции *wA(n)* и *wB(n)*, являющиеся квадратично зависимыми, в отличие от независимых *wL(n)*, *wR(n)*, формируемых полной композицией «11-set» (структурные формулы вычисления функций приведены в рамках на поле *рис. 5*). Как выглядят функции *wL(n)*, *wR(n)* для «*C*-овала 1» и «*C*-овала 2» и кривые *W*-отображений *wB(wA)*, *wR(wL)* для них, показано на *рисунке 6*.

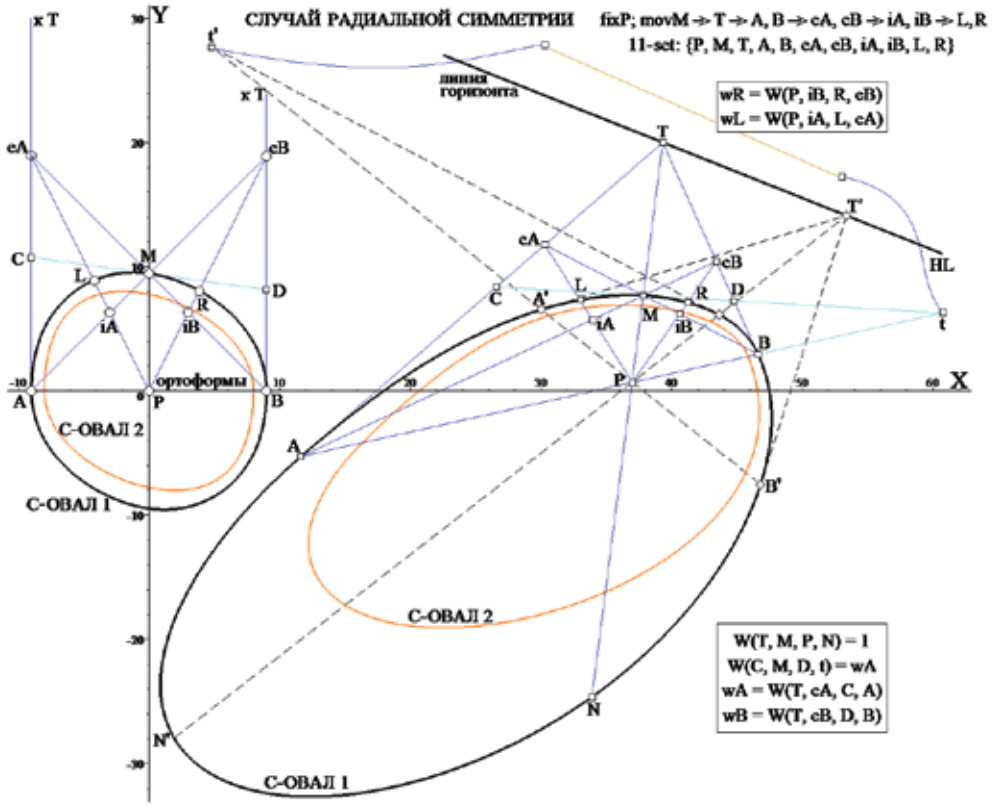


Рис. 5. Вид C-овалов и варианты построения проективно инвариантного базиса.

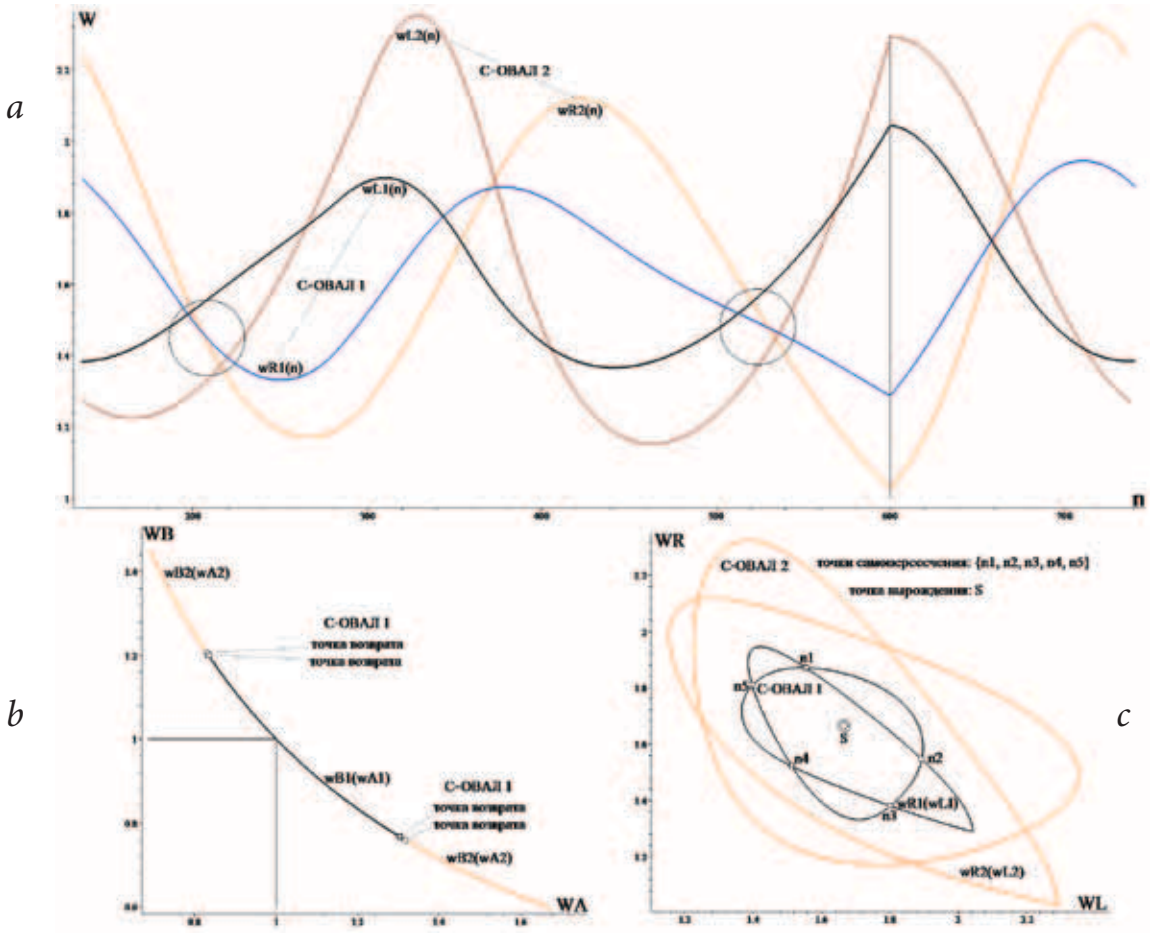


Рис. 6. Вид W-функций C-овалов в базисе «11-set» (a) и их W-отображений (b, c).

 DOI: 10.22204/2410-4639-2016-092-04-38-54 № 4 (92) октябрь–декабрь 2016 г.

Есть ли вообще какой-либо смысл в задачах технического распознавания привлекать 7-точечный шаблон, если носителем отображений *C*-овалов является «эталонная коника»? В ряде случаев это можно объявить полезным, так как координаты точек возврата на кривой *WB(WA)* многократного самопокрытия вполне «автографично» (инвариантно и индивидуально) характеризуют описываемый *C*-овал (на *рисунке 6b* метки ромб и квадрат позиционируют четыре точки возврата для *WB*1(*WA*1)). Универсальное описание *C*-овала (см. выше тезис в материале анализа *R*-овалов) обуславливает целесообразность привлечения «характерных» точек и интегральных свойств (площадь и топология петель) в обход вычислительных затрат на оценку метрики близости [7] *W*-кривых при классификации овалов в рамках проективной их эквивалентности. На *рисунке 6c* вполне выразительны и свойство уменьшения площади отображения для овала «1», «точнее» аппроксимируемого эллипсом, и заметные отличия в композиции точек самопересечения (на кривой *wR*1(*wL*1) сет пяти точек помечен литерами *n*1 – *n*5 при значках «квадрат»). Перейдем к обсуждению репрезентации *A*-овалов.

Проективно инвариантное описание *A*-овалов

В отличие от центрально симметричных кривых овалы с осевой симметрией обменивают «ключевые» полюс и поляру (как ЭС, где образ оси, являющийся уникальной плюккеровой полярой *A*-кривой, сменяет полюс образа центра *C*-овала), т.е. ЭС внешнего расположения становятся внутренними и наоборот. Это обстоятельство и детерминирует смену «ролей» точек и прямых в универсальном шаблоне: точек по-прежнему 11, но так как образ оси A-B *(рис. 7a)* фиксирован, а потому неподвижен и полюс *T* (в его качестве 0-мерного носителя *T*-образа *A*-овала), то все текущие смены позиций точек базиса должны осуществляться в границах неподвижного треугольника *A-T-B* при их определяющем положении «задающей точки *M*» (роль *M* не изменилась, она по-прежнему «сканирует» половину контура фигуры, но единственная фиксированная точка *C* у *C*-овала заменяется для *A*-кривой неподвижным треугольником *A-T-B*). В модельном примере для двух *A*-овалов на *рисунке 7a* показаны схема формирования универсального базиса (слева для ортоформы и справа в проекции) и вид их *W*-функций (*b*) и *W*-отображений (*c*). В качестве «методической справки» там же (*рис. 7a*, в рамке) приведены структуры вурф-композиций (*W(D, C, M, T)*, *W(A, iA, M, eB)* и *W(D, C, M, S))*, не подходящих на роль независимых вурф-функций базиса ввиду их «универсальной константности». Вид значительно различающихся *wR*1(*wL*1) и *wR*2(*wL*2) (ортоформы которых визуально едва различимы) говорит сам за себя. На этом завершим рассмотрение схем инвариантной репрезентации *A*-овалов, проективно инвариантное описание которых, как и для кривых *R*- и *C*-типов, можно строить посредством схемы «11-set», и перейдем к идее «адаптации» развитого подхода в анализе кривых, не имеющих каких-либо ЭС (для двух «спецзадач» распознавания).

Случаи: овал общего вида с фиксированной точкой и «Овал + *HL*»

Интересен вариант использования 11-точечного шаблона, предложенного нами для обработки *C*-овалов, в качестве инструмента проективно инвариантного описания овала общего вида (не обладающего скрытой симметрией), в жесткой композиции с которым фиксирована точка *P* внутреннего поля фигуры. Процедура фактически та же: *A* является задающей точкой базиса, а *P*, утратив роль образа центра симметрии, остается для *B* ее азимутальным ориетиром *(рис. 8a)*. У *L* и *R*, как и для *C*-овала, аналогичен вклад в оценку вурфов *W(eA, R, iA, P)* и *W(eB, L, iB, P)*, которая оказывается осуществимой благодаря вычислению позиции *M* (что делает возможной оценку позиций *eA, iA, eB и iB*). Отметим, что модель овала на *рисунке 8a* сгенерирована по формулам для R_3-кривой со слабо нелинейной трансформацией ее полярных радиусов, производимой из центра *P*. Этим обстоятельством объясняется осесимметричная форма *wB(wA)* (ее уменьшенный вид дан на *рисунке 8c*).

Если отнести этот случай (уже в качестве метафоры) к категории дуального образа анализируемой

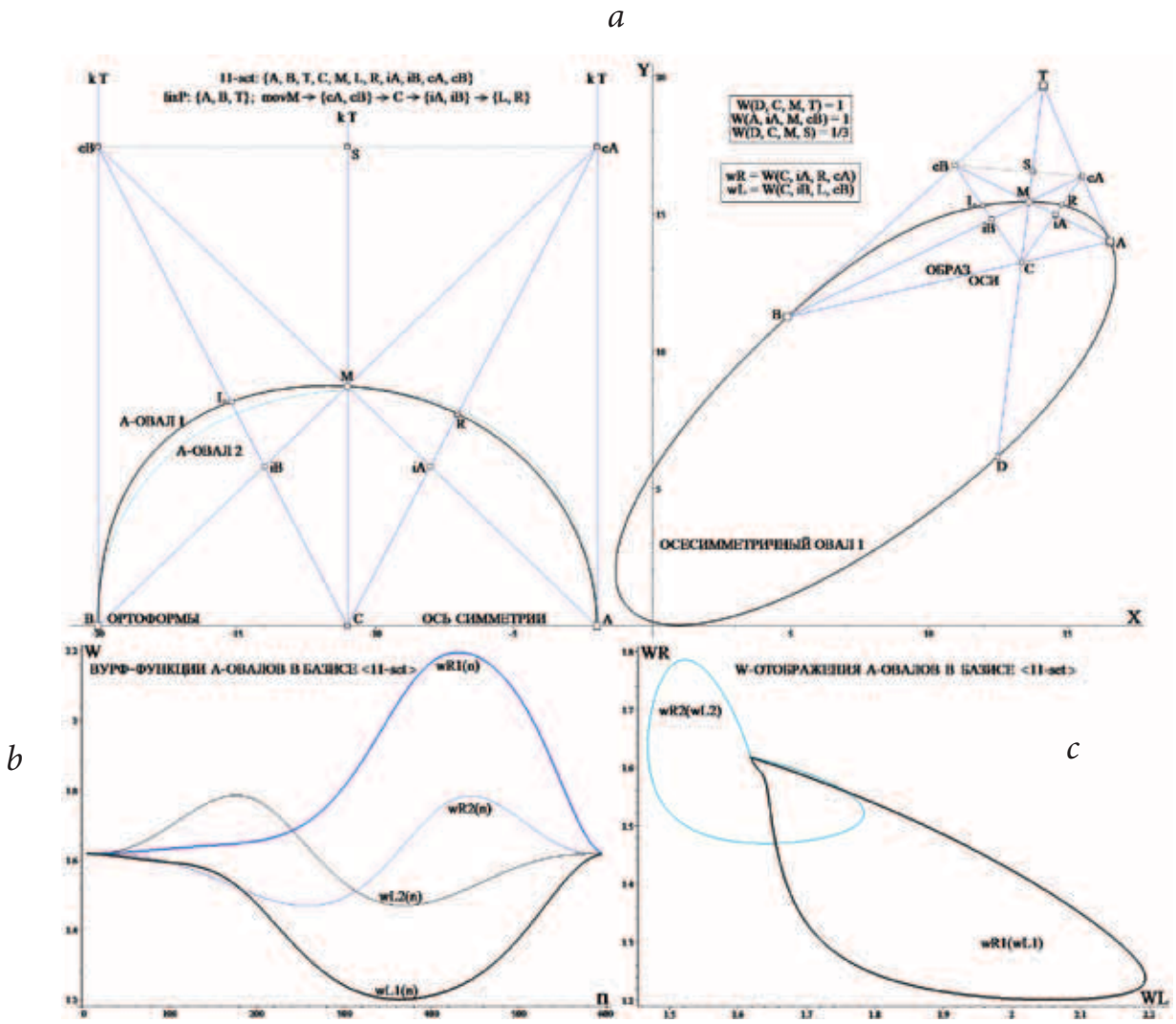


Рис. 7. Осесимметричные овалы (a): вид W-функций (b) и W-отображений в базисе «11-set» (c).

фигуры как кривой внешнего расположения, поставляющей базису координаты подвижной точки, задающей необходимые переменные вурф-функций, то эту роль следует связать с введенным нами ранее инструментом анализа овалов, названным *T*-полярой. Действительно, полюс *T* скользит (в итоге вращения поляры *A-B* вокруг *P*) вдоль кривой *T(n)* согласно формальному определению *T*-поляры [8]. Отметим также, что для объекта «Овал + *P*» схема «11-set» к двум независимым вурф-функциям может добавить третью *wM(n) = W(T(n), M(n), P, D(n))* (которая для *C*- и *A*-овалов была непригодной, являясь константно единичной), что расширяет информационную базу «атрибутов распознавания». Но и помимо перспектив технического использования этого подхода, схема «11-set» самоценна в качестве инструмента анализа объектов выделенного типа, обеспечивающего их проективно инвариантную репрезентацию. Не станем столь же подробно описывать «расклад ролей» в схеме базиса – применительно к семейству овалов с фиксированной внешней точкой достаточным для понимания будет поставить задачу «по аналогии». Так, при модификации схемы для фигур «Овал + *i*P» (от internal) прообразом ее был базис *C*-овала, в случае же объекта «Овал + *e*P» (от external) «роли точек в статике и динамике» следует ассоциировать со сценарием схемы, рассмотренной нами для *A*-овалов (прямую *A-B* «образа оси» должна заменить неподвижная плюккерова поляра полюса *eP*).

Завершим «адаптационный» прогноз приложений развитого нами концептуального подхода примером «специфически технического» его использования. Рассмотрим вариант системы оптического распознавания овалов, не применяющей для анализа объектов их центральных проекций, зарегистрированных при неизвестных для системы параметрах проецирования, но требующей обязательного наличия при каждой проекции

 DOI: 10.22204/2410-4639-2016-092-04-38-54 № 4 (92) октябрь–декабрь 2016 г.

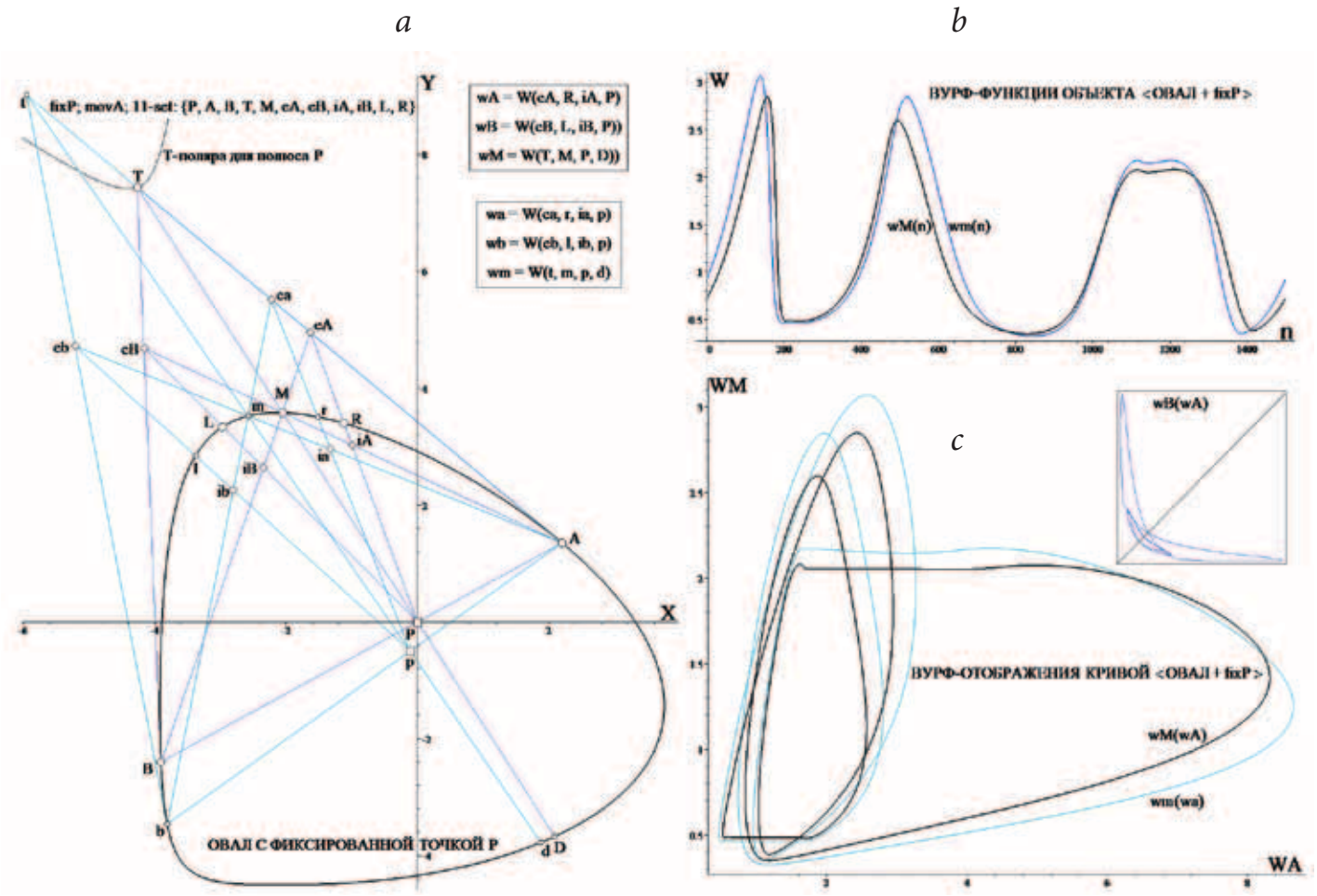


Рис. 8. *Базис овала с фиксированным внутренним полюсом P (a) и его W-продукты (b, c).*

объекта данных о 3D-ориентации оптической оси сенсора относительно плоскости анализируемой фигуры (с учетом знания фокусного параметра камеры). В рамках этого допущения каждая из входных картин может быть дополнена позицией *HL* – положением линии горизонта для этой конкретной копии фигуры. Описанная ситуация с некоторым усложнением схемы может быть сведена к сценарию для *C*-кривых, однако не слишком убедительной будет демонстрация «триумфа универсальной композиции», если стоящую задачу можно решить «на порядок проще». Для этой цели подойдет диспозиция поляр и полюсов, декларированная для коник теоремой взаимности [9]. Именно для них доказаны условия дуальной симметрии отношений, которые в нашем случае (т.е. при условии, что объект распознавания не принадлежит семейству квадратичных кривых) гарантируют результат, отличный от вырождения (когда «отображение описывает точка с координатами (1, 1)»). На *рисунке 9* показаны: *a* – схема формирования базиса (в смене синей композиции композицией коричневой – в динамике перехода от «ведущей точки» A к точке a), *b* – пример триады вурф-функций для овала, промоделированного для *рисунка 8*, и *c* – две кривых его *W*-отображения.

Заключение

Резюмируем главный итог рассмотренных модельных реализаций, иллюстрирующих выдвинутый на обсуждение тезис: в рамках концепции о целесо-образности введения понятия дуальный образ кривой (овала *SO*, имеющего ЭС) и его имплементации в задачах проективно инвариантного *W*-описания симметричных фигур – безотносительно к ракурсу оптической регистрации и типу симметрии объекта «*SO*» – предлагается вычислительная

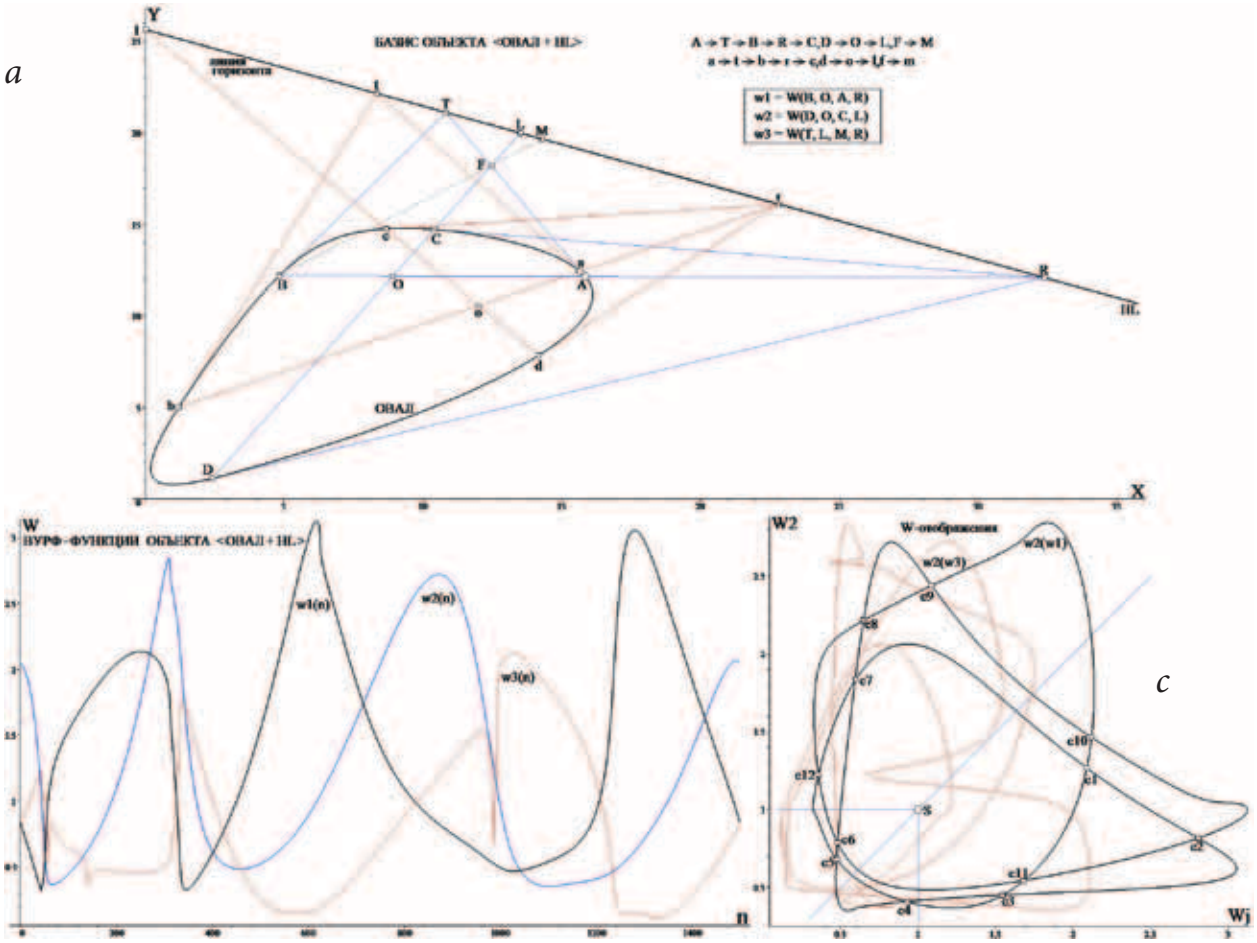


Рис. 9. *Схема получения базиса (шаг: от заглавных литер при метках композиции – к строчным) (a) и вид его продукций (три вурф-функции (b), W-отображения (c)).*

схема скользящего базиса «11-set», формирующая *W*-script *SO* на основе использования исключительно распределенных (интегральных) свойств *SO*. Компьютерная реализация тезиса проведена нами на модельных (аналитически генерируемых) объектах с машинной точностью представленными характеристиками их симметрии. Реалистичность задания свойств оптики сенсора также не входила в круг моделируемых проблем: закономерности трансформации образа (при смене ракурса) были аппроксимированы цент-ральной плоской проекцией (8-параметрическое преобразование плоскости в декартовом 3D-пространстве). Следует разграничить теоретический и прикладной аспекты проведенного исследования. В части возможного вклада в теорию (раздел «Проективно дифференциальная геометрия гладких выпуклых 2D-кривых»; [10-12]) представляется целесообразной попытка доложить научному сообществу о содержательности категории «дуальный образ *SO*» (в том числе и по причине ассоциативной ее связи с плюккеровским «дуальным образом алгебраической кривой» [13]). Структуру предложенного «универсального базиса 11-set» можно аттестовать некоторым «инструментальным развитием» понятия *плоскостной вурф*, введенного Депутатовым [6] (и развитого Глаголевым [4]). Попытка усмотреть общность в «поведении полюсов и поляр» у фигур разных типов (*R*, *C* и *A*) плоской симметрии, столь непохожих по своим проявлениям, оказалась в итоге плодотворной. Технический аспект настоящей работы интересен не набором программ, обеспечивших демонстрационный материал, а идеями, заложенными в схемы. И тут новые подходы могут оказаться востребованными (равно как и оригинальные постановки самих задач [3, 14]).

 DOI: 10.22204/2410-4639-2016-092-04-38-54 № 4 (92) октябрь–декабрь 2016 г.

Литература

1. ***П.П. Николаев***
Сенсорные системы, 2012, **26**(4), 280.
2. ***П.П. Николаев***
Сенсорные системы, 2013, **27**(1), 10.
3. ***П.П. Николаев***
Сенсорные системы, 2015, **29**(1), 28.
4. ***Н.А. Глаголев***
Проективная геометрия: Учеб. пособие, Москва, Высшая школа, 1963, 344 с.
5. ***Г.П. Акимова, Д.С. Богданов, П.А. Куратов***
Труды ИСА РАН, 2014, **64**(1), 75.
6. ***В.Н. Депутатов***
Матем. сб., 1926, **33**(1), 109.
7. ***H. Alt, M. Godau***
Int. J. Comput. Geom. Appl., 1995, **5**(1-2), 75.
DOI: 10.1142/S0218195995000064.
8. ***П.П. Николаев***
Сенсорные системы, 2011, **25**(3), 11.
9. ***П.С. Моденов***
Аналитическая геометрия, Москва, Наука, 1969, 699 с.
10. ***И.Ю. Овсиенко, С.Л. Табачников***
Проективная дифференциальная геометрия. Старое и новое: от производной Шварца до когомологий групп диффеоморфизмов, Москва, МЦНМО, 2008, 280 с.
11. ***С.П. Фиников***
Проективно-дифференциальная геометрия, Москва, УРСС, 2010, 264 с.
12. ***E. Cartan***
La Méthode de Repère Mobile, La Théorie des Groupes Continus, et Les Espaces Généralisés, Ser. Actualités Scientifiques et Industrielles, No. 194, Paris, Hermann, 1935, 65 pp.
13. ***А.А. Савелов***
Плоские кривые. Систематика, свойства, применения: Справочное руководство, Москва, Физматлит, 1960, 294 с.
14. ***P. Olver***
AAECC, 2001, **11**(5), 417. DOI: 10.1007/s002000000053.

English

A Projective Invariant Description of Ovals with Three Possible Symmetry Genera*

Petr P. Nikolaev –
A.A. Kharkevich Institute for Information Transmission Problems RAS
19-1, Bolshoy Karetny Per., Moscow, 127051, Russia
e-mail: nikol@iitp.ru

Abstract

This investigation deals with a conceptual approach to the problem of projectively invariant recognition of the planar curves and scheme of its numerical implementation that makes it possible to obtain projectively invariant descriptions of planar ovals with elements of the three genera of hidden symmetry – rotational, radial (central) and axial – not using for that purpose (as an element of the basis) a location of any contour point with a "special" projectively resistant characteristics. The author proposed to form a description relying upon some integral properties of the curve by means of secondary invariant structure called the tangential image of the oval. The latter structure can be computed in the end of the processing stage that estimates the location of the "image of the center of symmetry". Wherein earlier we had proposed, modeled and described the numerical search schemes for the coordinates of the center image. The author also discussed the application methods of the introduced analytical tools for the invariant description of ovals that do not have symmetry elements in the presence of additional point fixed outside the figure's outline.

Keywords: tangent line, orthoform, rotation index, Plucker's pole and polar line, wurf-mapping, tangent image.

* *The work was financially supported by RFBR (projects 13-01-12107 and 16-07-00836).*

Images & Tables

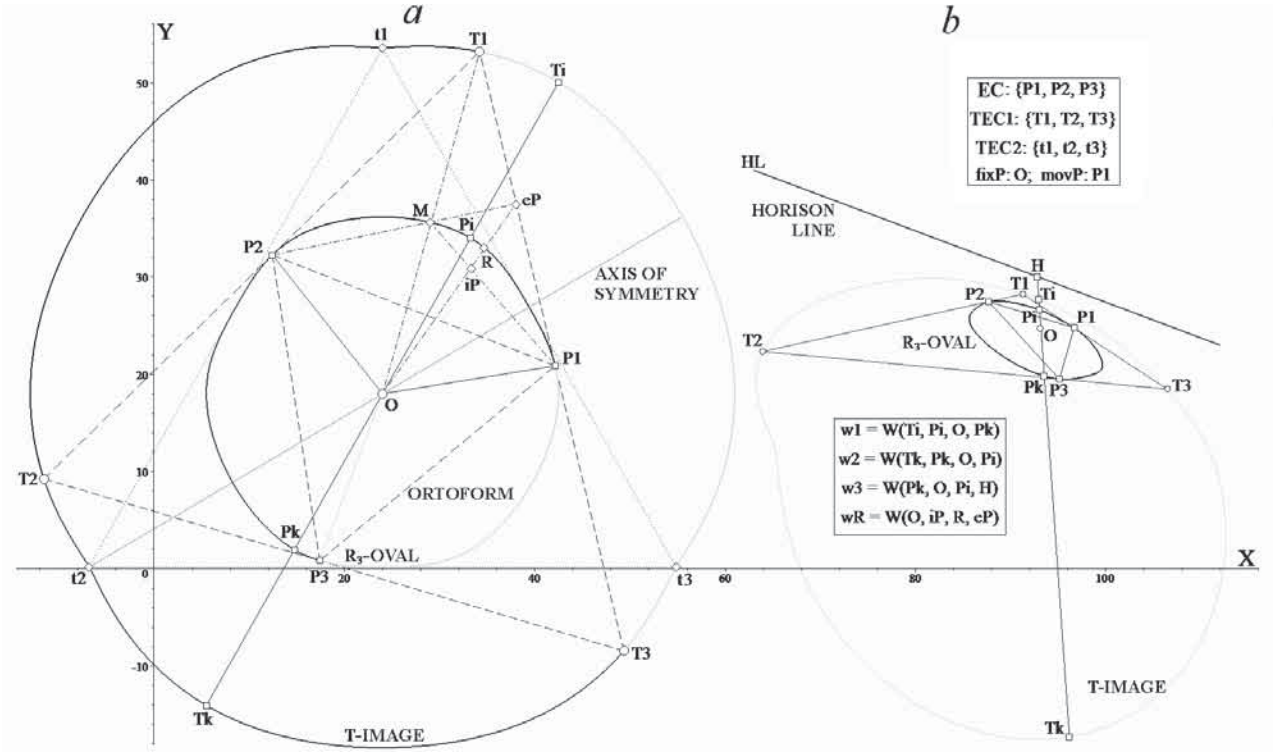


Fig. 1. The orthoform of R_3-oval and its T-image; the same in perspective; TEC - tangential ensemble of coordinates.

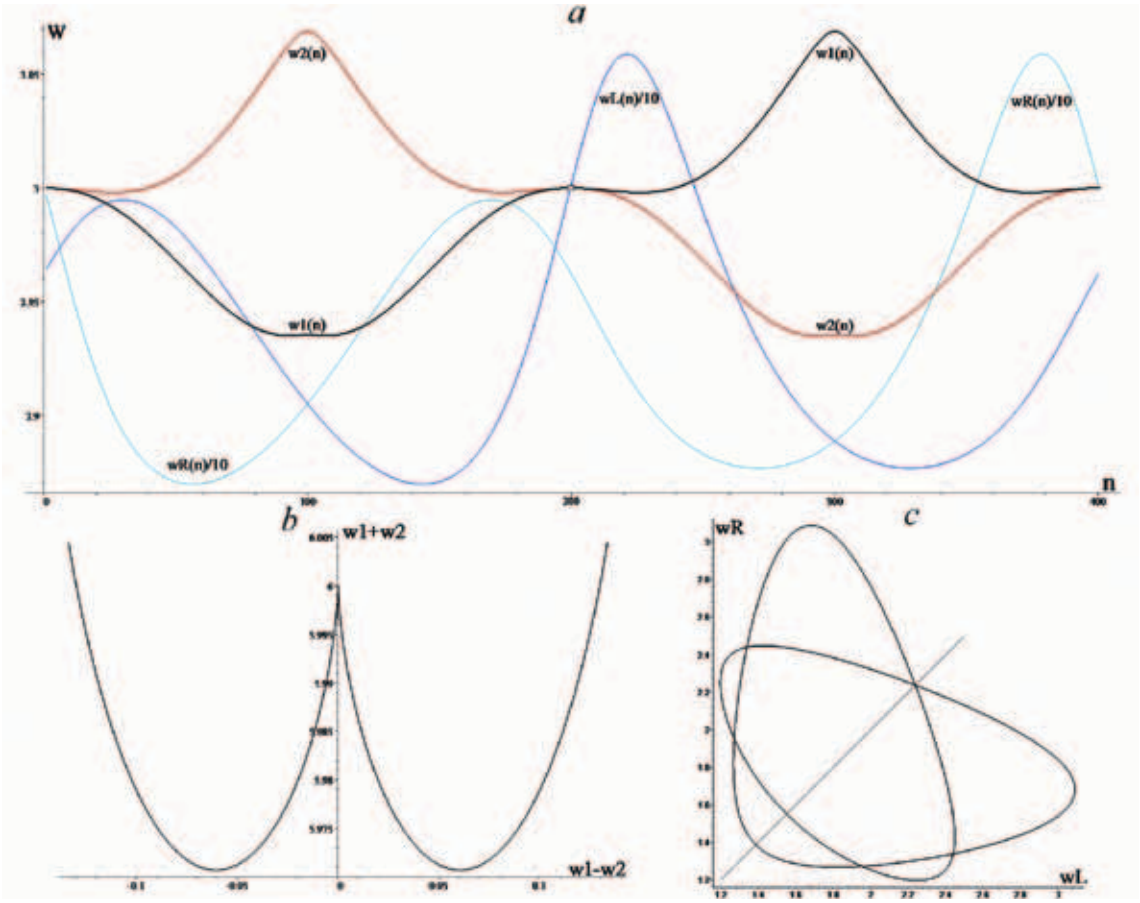


Fig. 2. The shape of wurf-functions (w1, w2, wL, wR) (a) and their W-mappings (b,c).

DOI: 10.22204/2410-4639-2016-092-04-38-54 № 4 (92) октябрь–декабрь 2016 г.

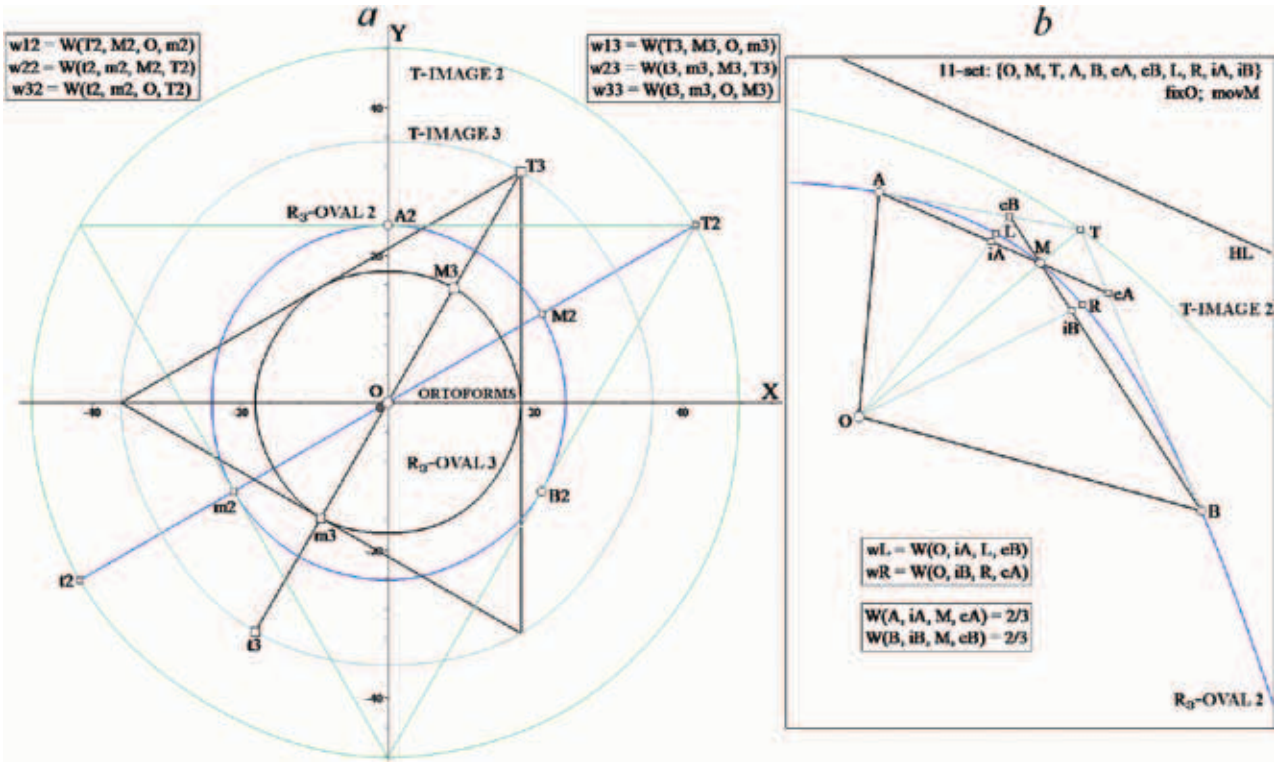


Fig. 3. R_3-ovals («2» and «3») in the orthoform (a) and the scheme of formation of the «11-set» basis (b).

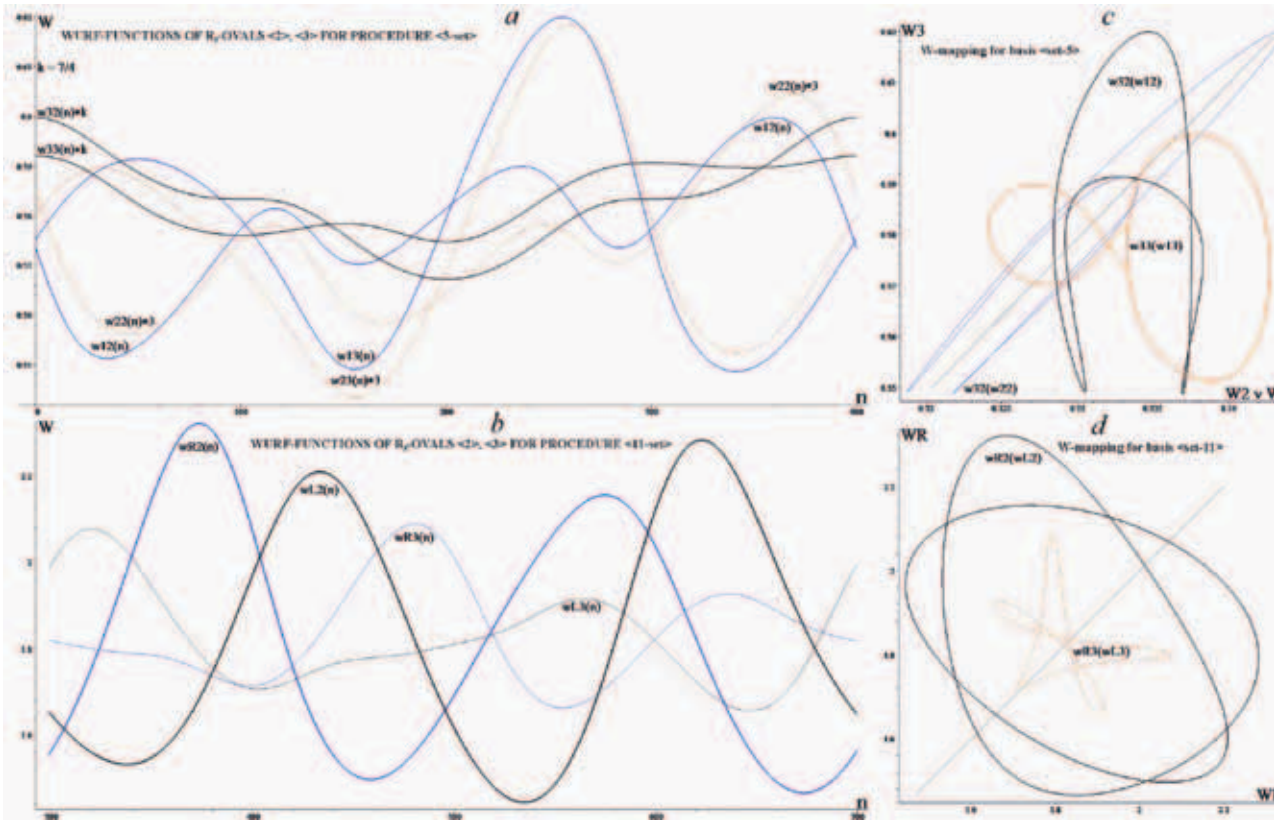


Fig. 4. The W-functions of R_3-ovals in the «5-set» and «11-set» bases (a, b) and the shape of their W-mappings (c, d).

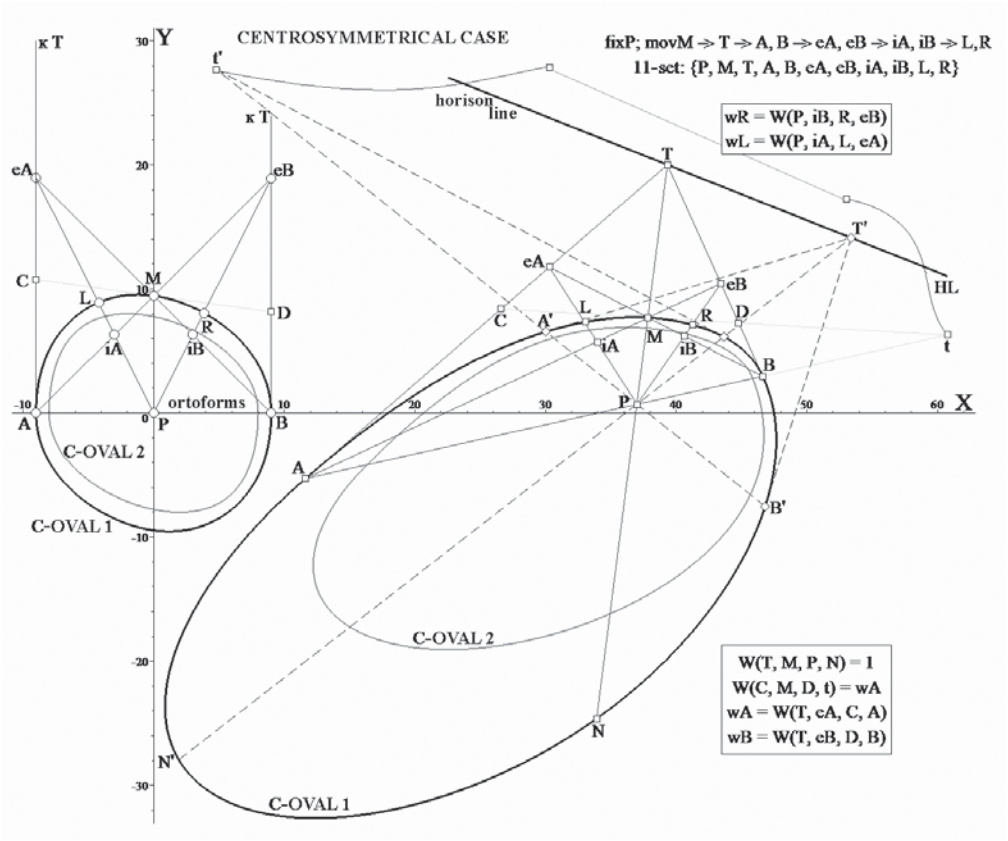


Fig. 5. The shape of C-ovals and variants of projectively-invariant bases construction.

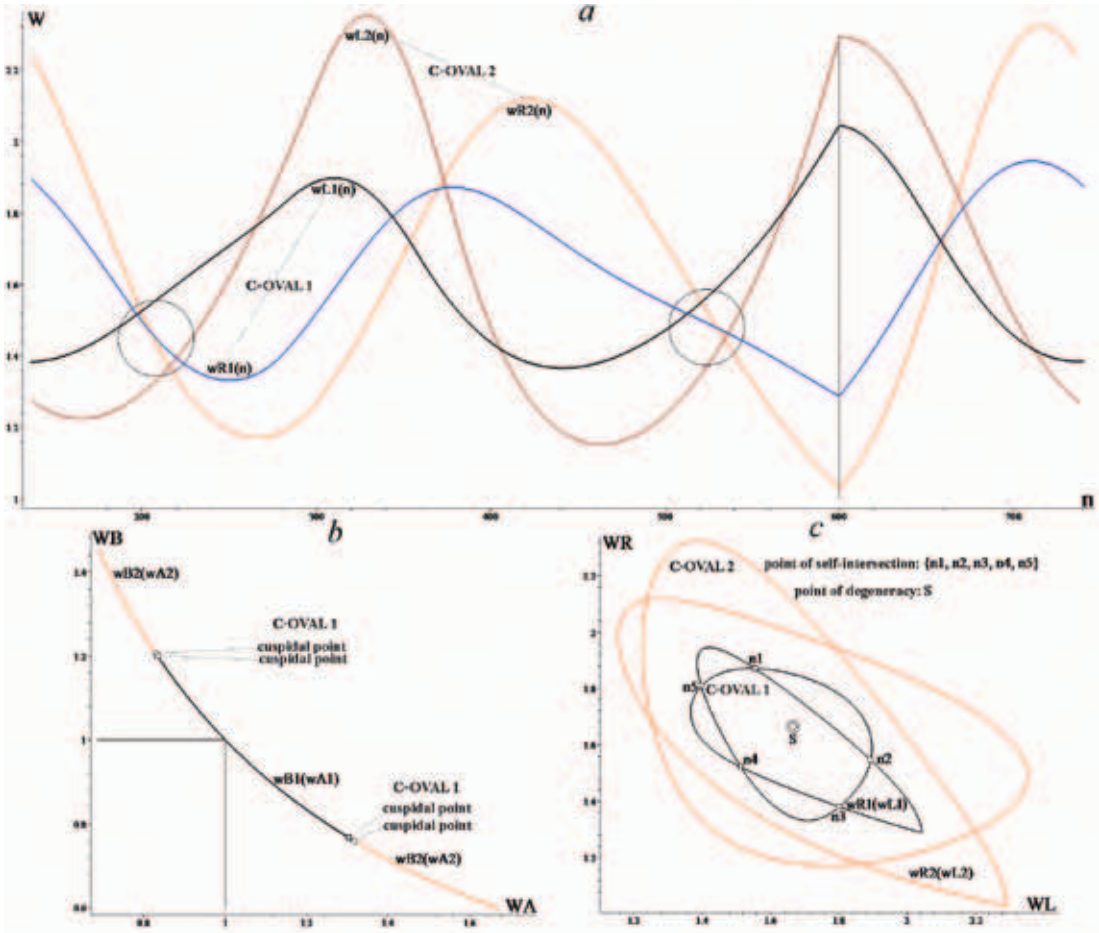


Fig. 6. The shape of the W-functions of C-ovals in the «11-set» basis (a) and their W-mappings (b, c).

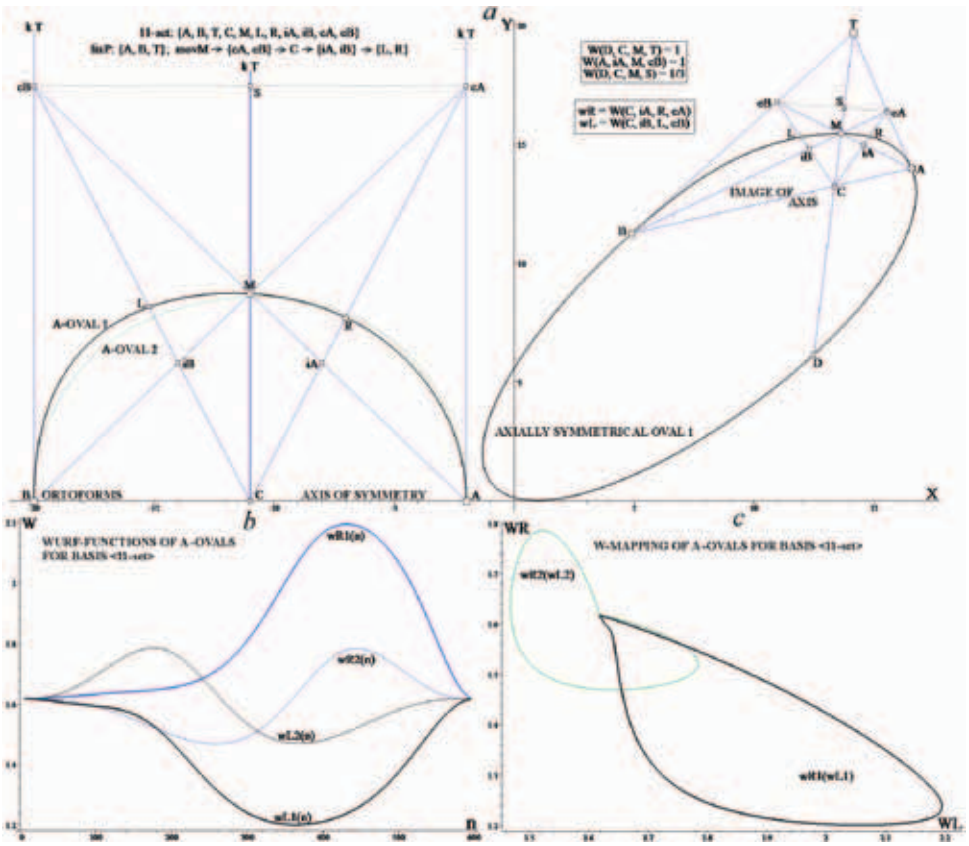


Fig. 7. Axisymmetric ovals (a): the shape of W-functions (b) and W-mappings (c) in the «11-set» basis.

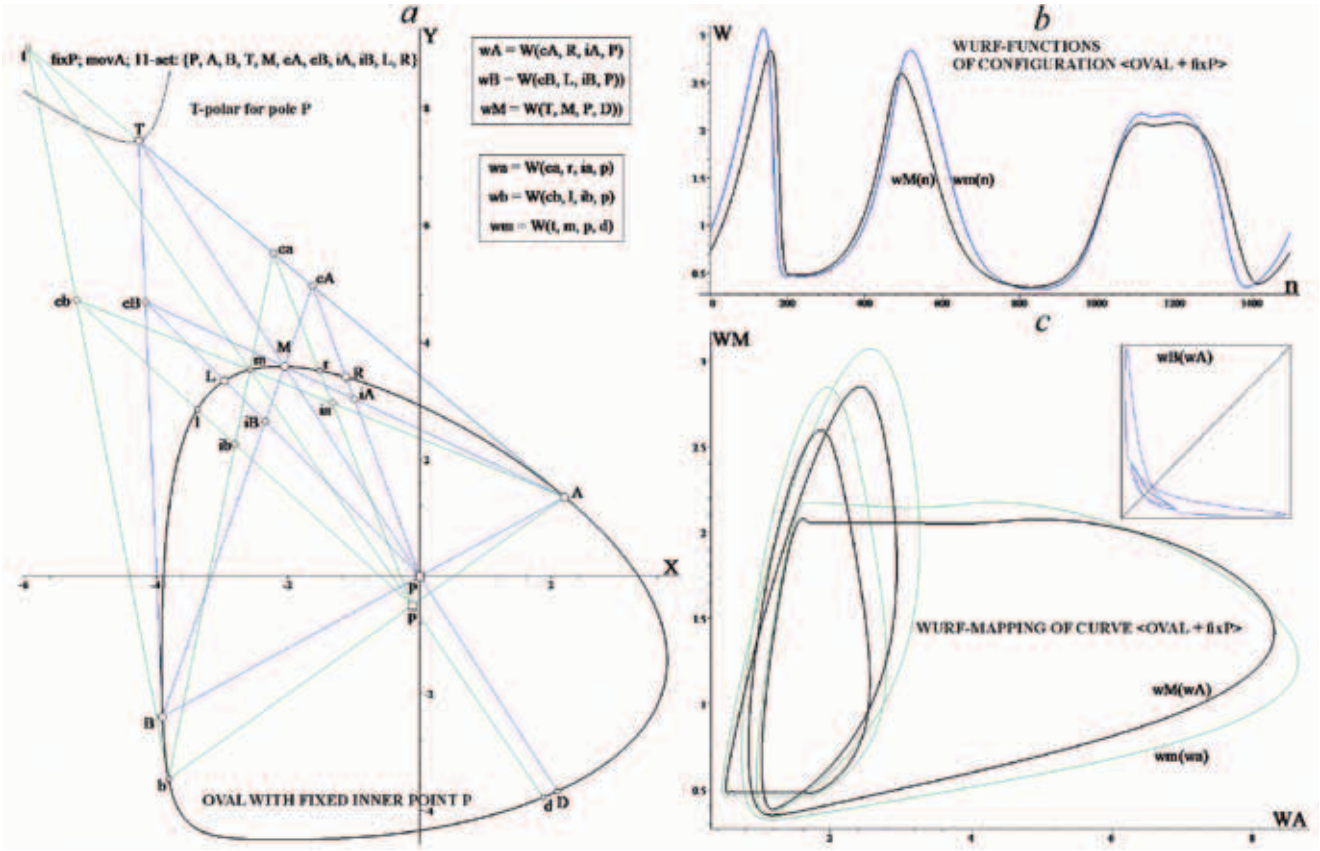


Fig. 8. The basis of the oval with a fixed internal pole P (a) and its W-products (b, c).

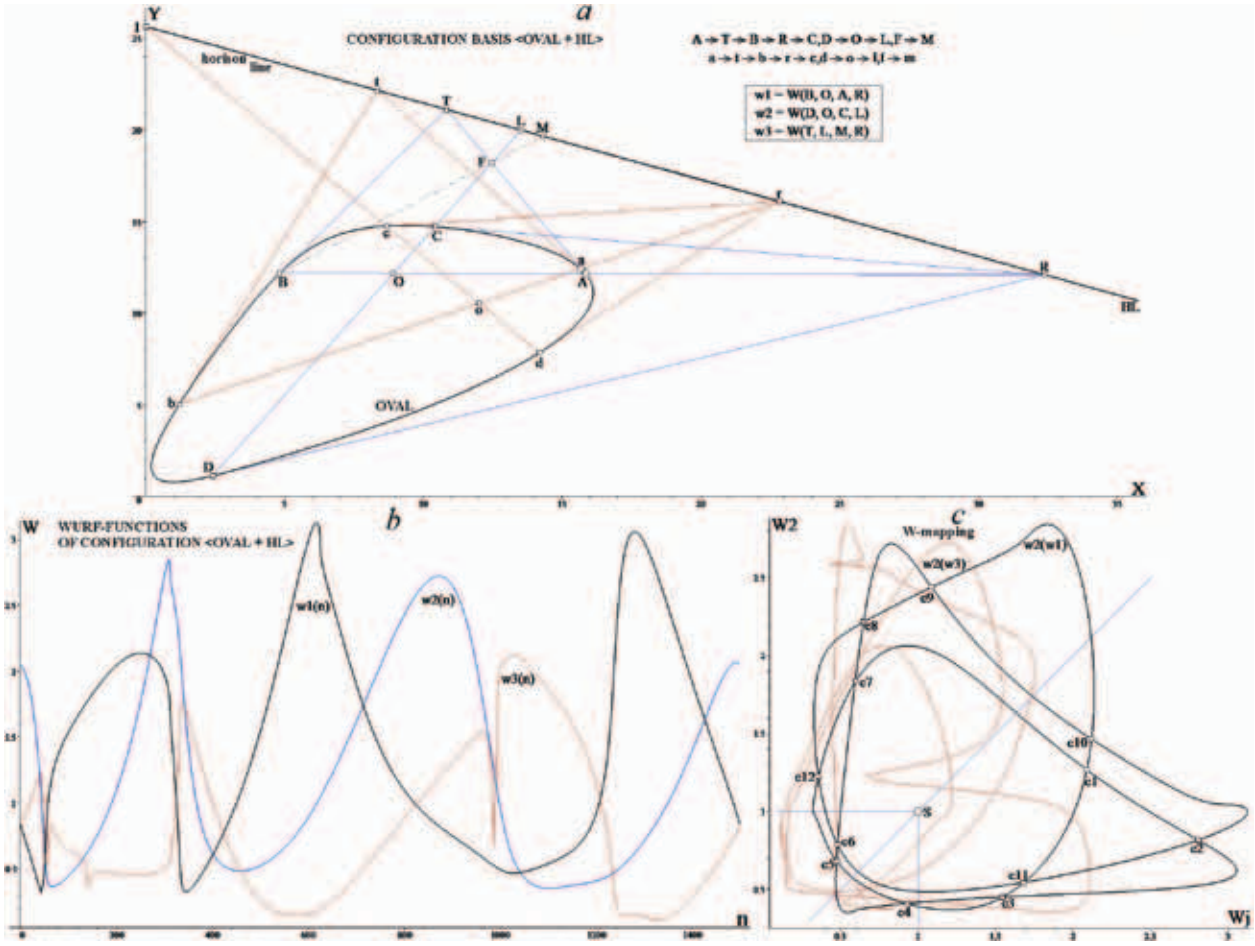


Fig. 9. The scheme of building the basis (a) of the «O + HL» object (step: from capital to lower-case letters in composition labels) and the shape of its products (b – three wurf-functions; c – W-mappings).

References

1. ***P.P. Nikolayev***
   Sensory Systems [Sensornye sistemy], 2012, **26**(4), 280 (in Russian).
2. ***P.P. Nikolayev***
   Sensory Systems [Sensornye sistemy], 2013, **27**(1), 10 (in Russian).
3. ***P.P. Nikolayev***
   Sensory Systems [Sensornye sistemy], 2015, **29**(1), 28 (in Russian).
4. ***N.A. Glagolev***
   Projective Geometry [Proektivnaya geometriya], RF, Moscow, Vysshaya shkola Publ., 1963, 344 pp. (in Russian).
5. ***G.P. Akimova, D.S. Bogdanov, P.A. Kuratov***
   Proc. ISA RAS [Trudy ISA RAN], 2014, **64**(1), 75 (in Russian).
6. ***V.N. Deputatov***
   Mat. Sb. [Mathem. proc.], 1926, **33**(1), 109 (in Russian).
7. ***H. Alt, M. Godau***
   Int. J. Comput. Geom. Appl., 1995, **5**(1-2), 75.
   DOI: 10.1142/S0218195995000064.
8. ***P.P. Nikolayev***
   Sensory Systems [Sensornye sistemy], 2011, **25**(3), 11 (in Russian).
9. ***P.S. Modenov***
   Analytic Geometry [Analiticheskaya geometriya], Moscow; Nauka Publ., 1969, 699 pp. (in Russian).
10. ***I.Yu. Ovsienko, S.L. Tabachnikov***
    Projective Differential Geometry. Old and New: from Schwartz Derivative to the Cohomology of Diffeomorphisms Groups [Proektivnaya differentsialnaya geometriya. Staroe i novoe: ot proizvodnoy Shvartsa do kogomologiy grupp diffeomorfizmov], Moscow, MTsNMO Publ., 2008, 280 pp. (in Russian).
11. ***S.P. Finikov***
    Projective Differential Geometry [Proektivno-differentsialnaya geometriya], Moscow, Editorial URSS, 2010, 264 pp. (in Russian).
12. ***E. Cartan***
    La Méthode de Repère Mobile, La Théorie des Groupes Continus, et Les Espaces Généralisés, Ser. Actualités Scientifiques et Industrielles, No. 194, Paris, Hermann, 1935, 65 pp.
13. ***A.A. Savelov***
    Planar Curves. Systematics, Properties, Applications: Guidebook [Ploskie krivye. Sistematika, svoystva, primeneniya: Spravochnoe rukovodstvo], Moscow, Physmathlit Publ., 1960, 294 pp. (in Russian).
14. ***P. Olver***
    AAECC, 2001, **11**(5), 417. DOI: 10.1007/s002000000053.

Исследование методов сегментации изображений текстовых блоков документов с помощью алгоритмов структурного анализа и машинного обучения *

Т.С. Чернов, Д.А. Ильин, П.В. Безматерных, И.А. Фараджев, С.М. Карпенко

Методы сегментации строки на символы являются важнейшими элементами при оптическом распознавании текста и образов документов. В работе рассматриваются два метода сегментации печатных текстовых полей. Первый метод развивает классические подходы к сегментации и содержит такие этапы, как анализ проекции, первичную простановку разрезов и динамическое программирование с учетом вероятностных оценок распознавания символов. Второй метод широко использует подходы машинного обучения, в частности сверточных и рекуррентных нейронных сетей, что делает возможным создание алгоритмов сегментации без разработки множественных эвристик, привязанных к конкретным типам полей документов, а также позволяет повысить устойчивость алгоритмов к различным искажениям, возникающим при съемке с мобильных устройств. В работе приводится сравнительный анализ двух методов на примере модуля сегментации системы распознавания паспорта гражданина Российской Федерации.

Ключевые слова: оптическое распознавание символов, анализ документов, сегментация, машинное обучение, анализ видеопотока.

** Работа выполнена при финансовой поддержке РФФИ (проекты №№ 13-07-12172, 13-07-12173 и 14-07-00730).*

Введение

Сегментация строки на символы является важнейшим этапом в процессе оптического распознавания символов (OCR), в частности при оптическом распознавании образов документов [1]. Сегментацией строки называется декомпозиция изображения, содержащего последовательность символов, на фрагменты, содержащие отдельные символы [2].

Важность сегментации обусловлена тем обстоятельством, что большинство современных эффективных алгоритмов оптического распознавания текста работает с отдельными символами, а не с целыми словами или строками. Ошибки неправильного проставления разрезов между символами часто являются причиной большинства ошибок конечного распознавания [2].

Поиск границ символов усложняется дефектами оцифровки (сканирования) документа, такими как рассыпание и склеивание образов различных символов [1, 3]. Кроме того, в случае использования стационарных или мобильных малоразмерных цифровых видеокамер в процессе распознавания документов возникает ряд других неблагоприятных особенностей формирования изображения [4]: дефокусировка или смазывание, проективное искажение, деформи-

ЧЕРНОВ
Тимофей Сергеевич
Национальный исследовательский технологический университет МИСиС

ИЛЬИН
Дмитрий Алексеевич
Институт системного анализа ФИЦ «Информатика и управление» РАН

БЕЗМАТЕРНЫХ
Павел Владимирович
ООО «Смарт Энджинс Сервис»

ФАРАДЖЕВ
Игорь Александрович
Институт системного анализа ФИЦ «Информатика и управление» РАН

КАРПЕНКО
Семен Михайлович
Институт проблем передачи информации им. А.А. Харкевича РАН

рование или изгиб документа. При съемке камерой в естественных сценах на изображениях часто возникают перепады яркости (тени, отражения, рефлексы), а также цветовые искажения и цифровой шум в результате низкой освещенности. На *рисунке 1* показаны примеры сложных случаев для методов сегментации полей паспорта РФ.

В работе приводится сравнительный анализ двух методов сегментации текстовых полей на примере модуля сегментации системы распознавания паспорта гражданина Российской Федерации. Особенностью системы является требование к распознаванию в реальном времени на мобильных устройствах, что сильно ограничивает круг возможных к применению методов сегментации по причине их низкой производительности и неустойчивости к неблагоприятным условиям съемки.

Первый рассматриваемый метод развивает классические подходы к сегментации и содержит такие этапы, как анализ проекции, первичную простановку разрезов и динамическое программирование с учетом оценок распознавания символов. Второй метод широко использует подходы машинного обучения, в частности сверточных и рекуррентных нейронных сетей, что делает возможным создание алгоритмов сегментации без разработки множественных эвристик, привязанных к конкретным типам полей документов, а также позволяет повысить устойчивость алгоритмов к различным искажениям, возникающим при съемке с мобильных устройств.

Существующие подходы к сегментации текста

Задача сегментации печатного текста исследуется на протяжении более четырех последних десятилетий [2, 5–9]. За это время было разработано множество различных методов сегментации как печатного, так и рукописного текста, сочетающих такие подходы, как анализ вертикальной проекции изображения строки [2], которое неустойчиво к шуму, сложному фону и склеиванию символов [2, 5]; анализ двумерных связанных регионов (компонент связности), состоящих из точек изображения, предположительно являющихся частью одного из символов [10]; постобработка, при которой с учетом модели языка и особенностей начертания символов в конкретном языке разбираются отдельные буквы, цифры, специальные символы, знаки пунктуации, математические знаки, применяются различные эвристики.

Так как сегментация строки проводится для ее последующего распознавания, возникли подходы второго типа, использующие результаты методов распознавания для поиска разрезов между символами [7, 11, 12]. Целью сегментации становится нахождение подмножества отрезков с минимальной суммарной штрафной оценкой распознавания [1], например с помощью динамического программирования [13].

В отдельный класс выделяются методы, вообще не использующие явную сегментацию на символы перед их непосредственным распознаванием [2]. Одним из многообещающих [14,15] подходов является распознавание текста с помощью

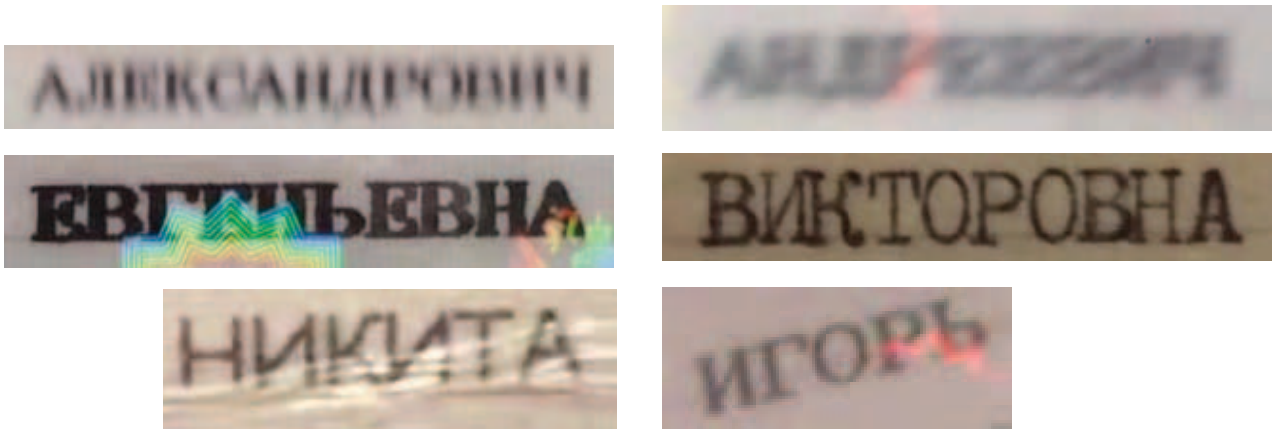


Рис. 1. *Изображения полей паспорта РФ с искажениями, усложняющими поиск границ символов, полученные в естественных сценах с помощью мобильной видеокамеры.*

DOI: 10.22204/2410-4639-2016-092-04-55-71 № 4 (92) октябрь–декабрь 2016 г.

искусственных нейронных сетей, имеющих рекуррентную архитектуру, например долгой краткосрочной памяти (Long Short-Term Memory, LSTM [16]).

Оценка качества методов сегментации

Целью сегментации текста на символы является его последующее распознавание, чем обусловливается популярность использования качества финального распознавания в качестве оценки качества алгоритма сегментации [17]. Оценкой же качества алгоритма распознавания может быть как точность распознавания отдельных символов или слов, так и среднее расстояние Левенштейна [18]. Показателями качества системы распознавания паспорта РФ в данной работе были положены точности распознавания каждого из полей документа (имени, фамилии, места рождения и т.д.) с точностью до символа по причине высокой стоимости единичной ошибки в отдельно взятом поле – ошибка даже в одном символе поля, идентифицирующего личность, является критической.

Дополнительными, не зависящими от конкретной модели распознавания, метриками качества сегментации, использованными в разработке, были среднее расстояние до ближайшего разреза, точность, полнота и *F1*-мера. Точность (precision, *P*) определяется как доля разрезов в результатах алгоритма сегментации; для них существует разрез из разметки, находящийся в 10%-ном диапазоне от средней ширины символа в поле. Полнота (recall, *R*) – наоборот, количество разрезов в разметке, которым найден соответствующий разрез в выходе алгоритма. *F1*-мера (*F1* score) – среднее гармоническое точности и полноты, вычисляемое по формуле $F1 = 2\frac{P \cdot R}{P + R}$.

Печатный текст в полях паспорта РФ может иметь разную ширину символов среди разных паспортов, но в рамках одного паспорта он, хотя и не является полностью моноширинным, часто имеет большую долю схожих по ширине символов. Таким образом, имеет смысл ввести среднюю ширину символов в конкретном поле конкретного документа разметки и использовать ее как нормировку результатов сегментации среди полей документов выборки.

Развитие классических подходов к сегментации строки

Опишем алгоритм сегментации строки, сочетающий классические подходы к сегментации текстовых блоков. Алгоритм реализован в качестве подсистемы сегментации изображения поля документа в системе распознавания паспорта гражданина РФ.

На вход подсистеме сегментации подается вырезанная зона паспортного поля в градациях серого, а также результат применения к данному изображению специального фильтра, подавляющего фон и повышающего контрастность изображения (примеры входа представлены на *рисунках 2* и *3*). Для входной зоны также определен тип поля – один из четырех заранее допустимых классов (common, digits, name, gender).

Рис. 2. *Вырезанная зона паспортного поля.*

Рис. 3. *Отфильтрованная зона паспортного поля.*

Каждый из указанных классов имеет свои особенности при обработке, которые будут описаны далее.

Обработка начинается с первичной оценки степени «моноширинности» шрифта, используемого в поле. Если он признается моноширинным, то запускается специализированная версия сегментатора, базирующаяся на идеях динамического программирования, для поиска набора оптимальных разрезов с фиксированной ожидаемой шириной символов. В ситуации, когда символы имеют различную ширину, запускается основной сегментирующий алгоритм. Вначале осуществляется поиск базовых линий текста, что позволяет сузить зону обработки. В случае работы с полем типа «Пол» применяется дополнительное уточняющее правило поиска опорных линий. Уточненная зона передается на так называемый

«первый проход сегментации» (результат сужения зоны интереса представлен на *рисунке 4*).

Рис. 4. *Уточнение зоны по базовым линиям.*

Проход сегментации стартует с простановки первичных разрезов: изображение проецируется на ось абсцисс таким образом, что в каждой точке оказывается минимальное значение яркости по соответствующему столбцу картинки. Каждую точку собранной проекции можно рассматривать как некоторую нормированную оценку правильности простановки разреза в ней. Чем светлее получившееся точка, тем вероятнее наличие разреза в данной позиции (в рассматриваемой фотометрике: 0 – черное, 255 – белое). Пример проекции приведен на *рисунке 5*.

Рис. 5. *Проекция изображения.*

Далее запускается рекурсивная процедура «дробления» проекции на сегменты. В идеале каждому знакоместу соответствует один сегмент, представляющий собой отрезок, задаваемый точкой на проекции и его шириной (в пикселях), а также содержащий оценки разрезов на своих краях. Процедура разбиения продолжается до тех пор, пока существуют сегменты, больше определенного размера (т.е. запрещается иметь в строке слишком широкие символы). Результат работы алгоритма приведен на *рисунке 6*. После данного этапа допустимо наличие неверных разрезов букв, а также их пропуск в местах «склеек» символов.

Рис. 6. *Первичная расстановка разрезов.*

После завершения работы алгоритма полученные сегменты передаются на процедуру «обжатия» символов. Данный этап преобразует найденные сегменты в коробки символов, плотно окаймляющих буквы, после чего запускается фильтрация компонент (результат показан на *рисунке 7*).

Рис. 7. *Результат обжатия компонент и зачистки мусора.*

Следующий этап является центральным во всем алгоритме. Для всех типов полей (кроме поля «Пол») запускается так называемая процедура клейки – разрезания символов. Каждое потенциальное знакоместо *cell* описывается своей прямоугольной зоной на исходном растре *zone*, качеством разреза слева и справа (cut_{left} и cut_{right} соответственно), а также результатом распознавания своего содержимого $recog_{result}$. Для распознавания растра, соответствующего коробке, используются специально обученные нейронные сети, содержащие в своем алфавите символ мусора.

Алгоритм выбора оптимального набора знакомест базируется на идее динамического программирования [19]. В качестве входных данных принимается множество *cells*, состоящее из знакомест, полученных после этапа чистки мусора. Оценкой отдельно взятого знакоместа *cell* будем называть результат вычисления выражения следующего вида:

$$score(cell) = recog_{conf} \cdot geom_{conf} \cdot cuts_{conf},$$

где $recog_{conf}$ – качество распознавания символа в прямоугольном растре знакоместа, $geom_{conf}$ – степень соответствия зоны символа полученным результатам распознавания, $cuts_{conf}$ – произведение оценок левого и правого разрезов знакоместа.

Обозначим через *width (cell)* ширину знакоместа. Из экспериментальных данных установлено, что ширина корректной зоны символа не может превышать значения *C*, заданного в пикселях. Для каждого знакоместа перебираются возможные его объединения с позициями слева (до тех пор, пока ширина получающегося после слияния знакоместа не становится больше указанного порога *C*). На основании вычисленных оценок выбирается или наилучший вариант объединения смежных коробок символов, или оставляется исходная зона знакоместа. Данные оценки заносятся в массив с именем

по следующему правилу:

Обратный проход по данному массиву позволяет восстановить последовательность знакомест с общим наилучшим качеством. Необходимо отметить, что на данном этапе активно задействуются различные эвристики, зависящие от типа обрабатываемого поля, а также получаемых результатов распознавания. После выбора знакомест осуществляется «финальная» обработка, зависящая от типа поля, и высчитывается итоговое качество получившейся нарезки на символы. Если оно признается удовлетворительным, то подсистема заканчивает свою работу *(рис. 8)*.

В обратном случае запускается «второй проход», отличающийся от первого только исходным набором точек разрезания. К данным, собранным на предыдущем этапе, добавляются вертикальные разрезы, призванные решить следующие проблемы: отсутствие естественного зазора между буквами из-за кернинга шрифта *(рис. 9, слева)*, а также слипание двух и более символов, что особенно часто встречается в случае применения шрифтов с засечками *(рис. 9)*. Анализируя форму штрихов символов как в верхних и нижних половинках широких знакомест, так и в центральной области, можно диагностировать данные случаи и добавить соответствующие разрезы в исходное множество альтернатив. Пример результата работы второго прохода приведен на *рисунке 10*.

Отметим, что поле «Пол» обрабатывается особым образом. Из полученных в результате обжатия знакомест удаляются небольшие компоненты (обычно это точки после МУЖ/ЖЕН или куски статических текстов паспорта), после чего выбирается наиболее близкая к центру растра коробка. Далее перебираются все компоненты справа и слева, и если расстояние до ближайшей не превосходит заданного порога T, то проводится слияние двух данных компонент. Процедура повторяется

$$dp[0] = 1.0$$

$$\left\{ dp[i+1] = \max_{(0 \le j \le i) \&\& (width(cell) \le C)} \{ dp[i+1] \cdot score(cell) \} \right\},$$

$$i = 0 \ldots n-1$$

Рис. 8. *Результат работы подсистемы.*

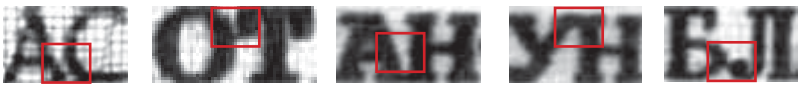


Рис. 9. *Примеры проблемных знакомест.*

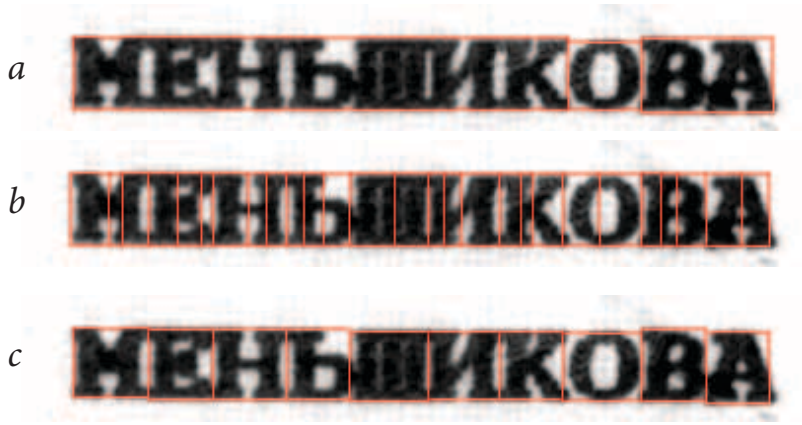


Рис. 10. *Пример результатов работы второго прохода: a – исходные разрезы, b – результат объединения двух групп разрезов, c – итоговый результат сегментации.*

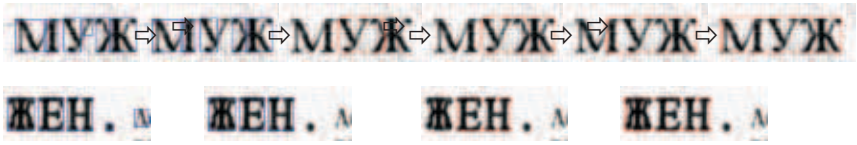


Рис. 11. *Примеры процесса обработки поля «Пол».*

до тех пор, пока есть возможность осуществить операцию слияния. Получившийся «иероглиф» является выходом подсистемы сегментации. Процесс формирования результата представлен на *рисунке 11*. Распознавание полученной компоненты осуществляется специально обученным классификатором, не требующим разбиения на отдельные символы.

Сегментация строки целиком с помощью методов машинного обучения

Как было показано ранее [1, 7, 11–13], в том числе в предыдущей главе, методы машинного обучения широко используются в современных алгоритмах сегментации. Однако их использование совмеща-

ется с дополнительными алгоритмами, такими как генерация первичных разрезов для обучаемой модели распознавания символов или динамическое программирование на выходных оценках этой модели.

Интерес представляет разработка метода сегментации, применяющего методы машинного обучения от начала до конца анализа изображения строки практически без дополнительной предварительной и последующей обработок (англ. end-to-end). Такие подходы отличаются тем, что не требуют тонкой ручной настройки под конкретный случай (шрифт, вид поля, тип документа), а требуют репрезентативную размеченную обучающую выборку достаточно большого размера. Это делает возможным упрощение и ускорение создания алгоритмов сегментации для новых типов полей документов, а также повышение точности и устойчивости к различным искажениям, возникающим при съемке. В качестве метода машинного обучения в работе были использованы искусственные нейронные сети различных архитектур, которые будут описаны далее вместе с методологиями составления обучающей выборки, идеальной разметки и функции потерь.

Подготовка и расширение обучающей выборки. Обучающая выборка содержит изображения выбранных полей паспорта РФ, обрезанных по базовым линиям текста, с размеченными позициями идеальных разрезов между символами. Разметка может проводиться как полностью вручную, так и полуавтоматически – существующий алгоритм сегментации проставляет разрезы, которые впоследствии проверяются и при необходимости корректируются людьми.

Подготовка выборки паспортов РФ и ее разметка – дорогостоящая операция, поэтому составление выборки достаточного объема для обучения аппроксиматора высокой точности, устойчивого к условиям съемки и ошибкам наведения полей, проблематично. С целью повышения устойчивости используется расширение (аугментация) обучающей выборки с помощью синтеза данных [20]. Синтез каждого образца осуществляется путем применения случайного набора преобразований, моделирующих трансформацию изображения поля. *Таблица 1* содержит иллюстрации описанных преобразований.

Формат разметки и выходных векторов сети. Вне зависимости от типа используемого универсального аппроксиматора для его обучения и работы необходимо определить функцию потерь (ошибки), которая будет минимизирована на обучающей выборке на стадии обучения параметров модели.

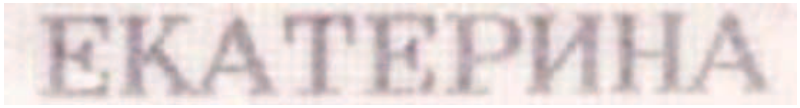
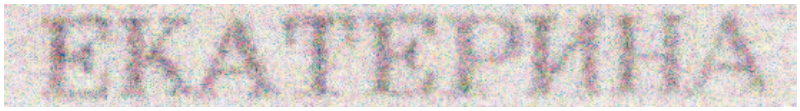
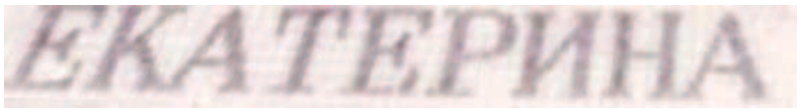
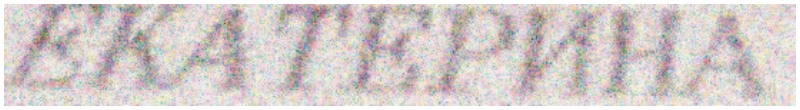
Требуются такие форматы разметки и выходов сети, которые поддерживают выставление любого допустимого количества разрезов, причем соответствующая им функция потерь остается пригодной для применения в обучении нейронной сети градиентными методами. Предлагается использовать следующую модель: вместо списка координат разрезов будем рассматривать вещественные вероятностные оценки нахождения разреза в каждом из столбцов пикселей изображения. Разметка разрезов тогда будет выглядеть так: нули во всех позициях, кроме позиций разрезов, в которых стоят единицы. Среднеквадратичное отклонение в таком случае также подходит в качестве функции потерь, но вычисляется уже между векторами оценок вероятностей. Итоговые позиции разрезов на выходе алгоритма получаются с помощью преобразования выходных вероятностных оценок, которое будет детально описано ниже.

Небольшие колебания относительно разрезов из разметки обычно не сильно влияют на качество распознавания, особенно если алгоритм сегментации ставит разрезы не непосредственно на границах символов (так, что между двумя символами стоит два разреза), а в области посередине (между символами стоит один разрез). Предложенная выше функция потерь в таком случае будет одинаково штрафовать выходные вероятности на позициях, не соответствующих идеальным разрезам, вне зависимости от их удаленности. Поэтому для смягчения штрафа при небольших колебаниях выхода сети предлагается использовать Гауссово размытие разметки с радиусом, пропорциональным средней ширине символа в данном изображении из тестовой выборки.

Архитектуры используемых сетей. Одной из наиболее популярных архитектур нейронных сетей в задачах анализа изображений является архитектура глубоких сверточных

 DOI: 10.22204/2410-4639-2016-092-04-55-71 № 4 (92) октябрь–декабрь 2016 г.

Таблица 1. *Преобразования изображений полей для расширения обучающей выборки*

Преобразование	Иллюстрация
Оригинальное изображение	
Гауссовый шум	
Проективные искажения	
Гауссово размытие	
Сдвиги	
Перемешивание букв	
Отражения	
Растяжение	
Комбинация преобразований	

сетей [21], которая была выбрана для первых содержательных экспериментов в создании обучаемого метода сегментации. Классическая модель сверточной сети состоит из нескольких сверточных слоев, составляющих карты признаков путем применения операции свертки с обучаемым ядром, чередующихся с субдискретизацией с целью уменьшения размерности карты признаков. Последние слои (заканчивающиеся выходным слоем) имеют полносвязную архитектуру.

Для используемых в работе нейронных сетей входными данными (векторами признаков) служат полутоновые растровые изображения полей паспорта РФ, приведенные к фиксированной ширине и высоте, например 200 × 20 пикселей. Размер выходного слоя, возвращающего вероятностные оценки наличия разреза в столбцах изображения, соответственно, также фиксирован и составляет 200 пикселей.

Скрытая часть нейронной сети состоит из сверточных слоев, за которыми следует два полносвязных слоя. За каждым сверточным слоем следует слой субдискретизации, а за каждым слоем субдискретизации и полносвязным слоем – эвристический слой случайного частичного обнуления (dropout). В качестве функций активации использовался гиперболический тангенс. В экспериментах также проверялась функция активации ReLU, с которой процесс обучения проходил значительно быстрее, но это приводило к сильному переобучению. На *рисунке 12* изображена схема

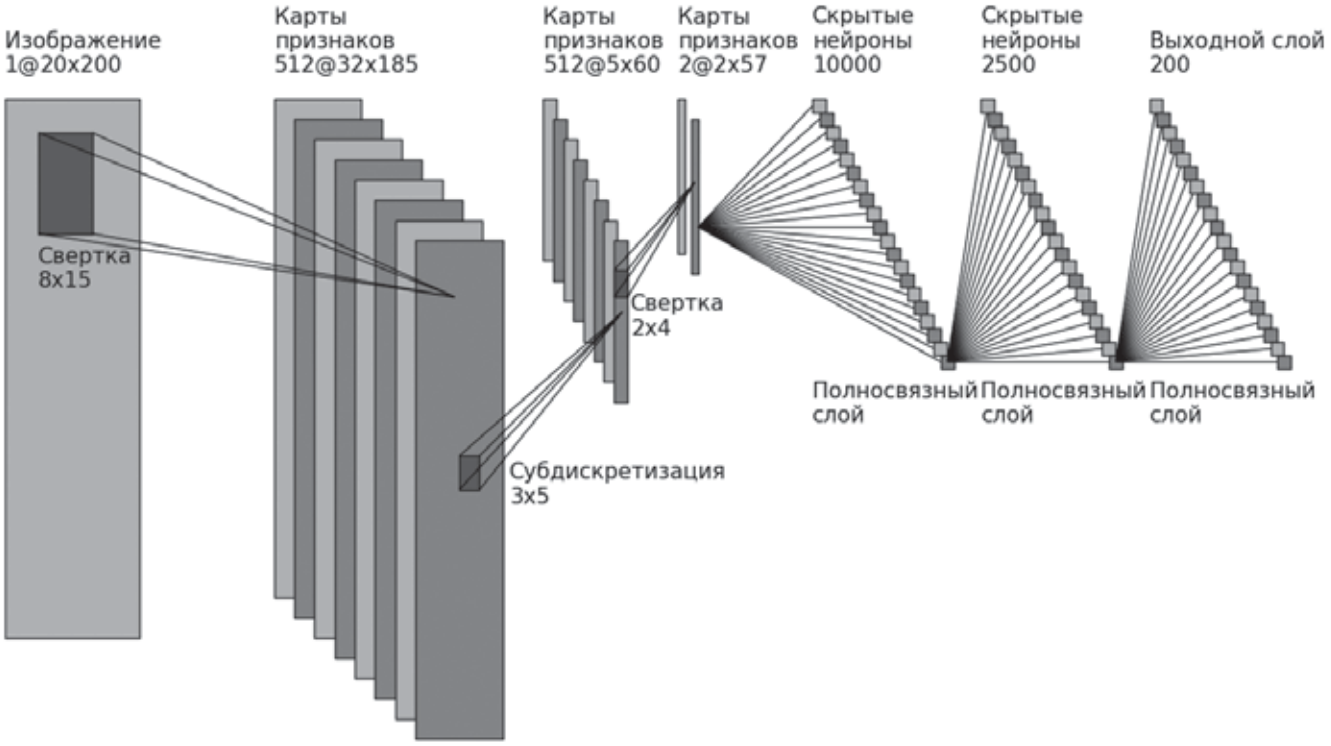


Рис. 12. *Схема сверточной сети, возвращающей вероятностные оценки разрезов для каждого столбца входного изображения.*

использованной в работе сверточной нейронной сети.

Обученная сегментационная сеть описанной архитектуры показала многообещающие результаты распознавания, о чем будет подробно рассказано в экспериментальной части работы, но они, тем не менее, были хуже, чем у инженерного подхода. С целью увеличения точности был применен метод последующей обработки выходов сверточной сети путем подачи их на вход дополнительной двусторонней рекуррентной сети.

Рекуррентные нейронные сети разработаны специально для работы с последовательностями: на вход помимо следующего элемента последовательности они принимают свой предыдущий выход, поддерживая некоторое скрытое состояние. В работе были использованы рекуррентные нейронные сети архитектуры долгой краткосрочной памяти (LSTM, [16]), зарекомендовавшие себя при частом применении анализа последовательностей, таких как распознавание печатного и рукописного текстов, речи и других. Сети LSTM архитектуры способны «запоминать» структуру последовательности; в случае сегментации строки под структурой можно понимать, например, среднюю ширину символов в строке и расстояние между ними.

Используемая двусторонняя рекуррентная LSTM-сеть принимает на вход последовательность, составленную путем приложения скользящего окна фиксированного размера (например 10) к выходному вектору вероятностных оценок сверточной сети. Таким образом, на каждой позиции входного вектора рекуррентной сети содержится последние 10 выходов сверточной. Двусторонняя направленность сети заключается в создании двух односторонних сетей, одна из которых обрабатывает последовательность слева направо, а другая – справа налево. Затем выходные векторы обеих сетей, соответствующие одинаковым позициям исходной последовательности, конкатенируются и передаются на вход полносвязному слою, после которого следует финальный слой, возвращающий аналогичные финальные вероятностные оценки. Важно отметить, что выходы сверточной сети, на которых проводится обучение рекуррентной, вычисляются заранее. Таким образом, сверточная сеть не меняется при обучении рекуррентной, что сильно ускоряет обучение. *Рисунок 13* содержит схему использованной

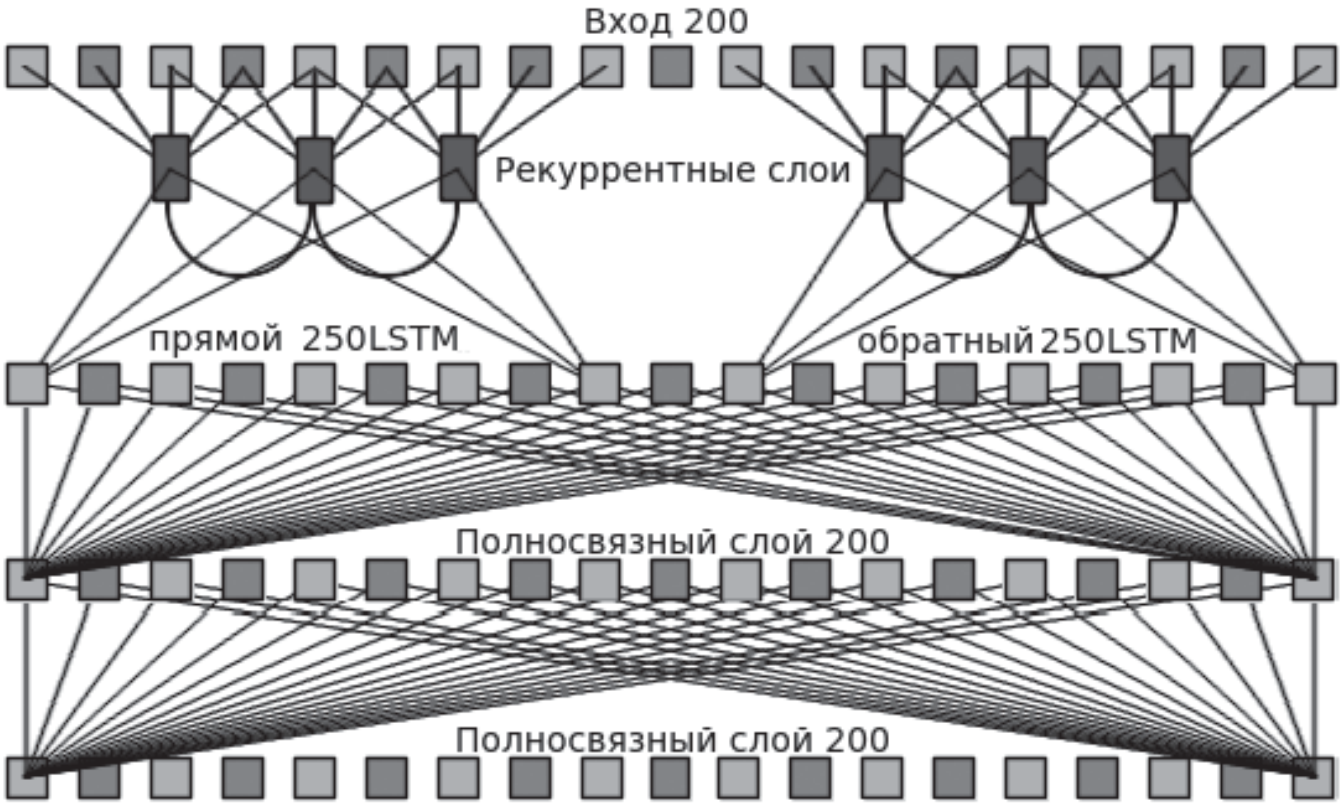


Рис. 13. Архитектура рекуррентной нейронной сети, входами которой являются выходные вероятностные оценки сверточной нейронной сети.

архитектуры рекуррентной сети на выходах сверточной.

После добавления рекуррентной сети на выходах сверточной общее финальное качество распознавания стало выше, чем у текущего «инженерного» подхода. Функцией активации в LSTM-сети также был гиперболический тангенс.

В качестве дополнительного исследования возможностей рекуррентных нейронных сетей в задаче сегментации строки был рассмотрен еще один подход. Его отличие состоит в том, что обучение сверточных и двунаправленных рекуррентных LSTM-слоев происходит в течение одного процесса обучения. Векторы признаков для сети составляются из регионов изображения некоторой ширины, сопоставимой с шириной символа, получаемых в результате приложения плавающего окна. Каждый регион изображения независимо проходит через сверточные и полносвязные слои, а затем их выходы конкатенируются и передаются в двунаправленный рекуррентный LSTM-слой аналогично предыдущей модели. Применение плавающего окна на исходном изображении позволяет поддерживать изображения любой ширины, приводя к фиксированному значению только высоту. Метод показал неплохие экспериментальные результаты, но, к сожалению, требовал намного больше времени на обучение, из-за чего приоритет был отдан предыдущему методу.

Преобразование вероятностных оценок в финальные разрезы. Для получения итоговых позиций разрезов на выходе алгоритма требуется выполнить преобразование вероятностных оценок. Простая пороговая фильтрация в данном случае не подойдет, поскольку выходные оценки сети сконцентрированы в большом количестве вокруг предполагаемых разрезов. Так как нейронная сеть аппроксимировала разметку, подвергнутую Гауссовому размытию вокруг позиций разрезов, довольно устойчивым и простым способом преобразования вероятностей в разрезы может быть фильтрация, оставляющая только локальные максимумы оценок, следующая после порогового отсечения с низким порогом. С целью устранения шу-

мовых срабатываний используется дополнительное Гауссово размытие, которое не меняет позиции сильных максимумов.

Метод фильтрации локальных максимумов является простым и показал хорошие результаты, но мы решили проверить, можно ли полностью избавиться от «инженерного» подхода на этапе преобразования вероятностей. В целях эксперимента была обучена еще одна сеть полносвязной архитектуры с небольшим количеством обучаемых весов, которая принимает на вход финальные вероятностные выходы итоговой сети, а на выходе возвращает аналогичные вероятностные оценки. Отличие состоит в том, что обучается она на оригинальных отметках разрезов, не подвергавшихся Гауссовому размытию. Вероятностные оценки на выходе последней сети уже подвергаются простой пороговой фильтрации без дополнительной обработки для получения финальных позиций разрезов. На *рисунке 14* изображены примеры работы описанного обучаемого алгоритма сегментации. Красным фоном отмечены вероятностные оценки на выходах сверточной сети, желтым – следующей за ней рекуррентной. Зеленый цвет обозначает вероятностные оценки рекуррентной сети, отфильтрованные по фиксированному порогу, синий цвет показывает оставшиеся разрезы – фильтрованные оценки, являющиеся локальными максимумами.

Эксперименты и результаты

Основными полями паспорта РФ в проведенных экспериментах были поля фамилии, имени и отчества. Размер исходной обучающей выборки составил 6000 изображений, после ее расширения с помощью синтеза данных – 150000 изображений. Размер тестовой выборки для оценки сегментации дополнительными метриками без распознавания – 630 изображений.

Тестовая выборка для распознавания полей содержала 1300 изображений паспортов РФ, по одному изображению каждого поля на документе. При распознавании полей на результатах сегментации всеми описанными методами использовалась одинаковая система распознавания паспортов РФ, не подвергшаяся изменениям. Важно отметить, что разработка системы распознавания велась к моменту проведения экспериментов достаточно долгое время вместе с разработкой первого алгоритма сегментации. Кроме

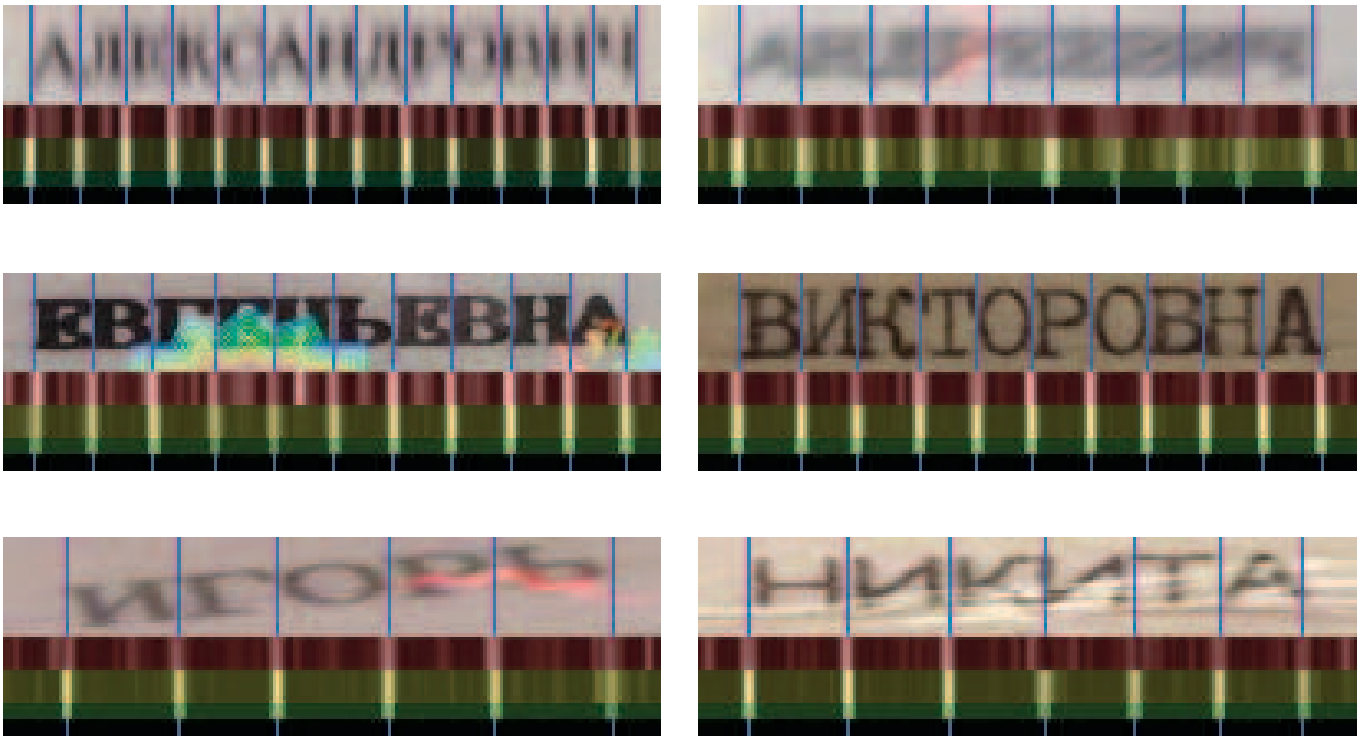


Рис. 14. *Пример работы обучаемого метода сегментации с выводом промежуточных результатов.*

DOI: 10.22204/2410-4639-2016-092-04-55-71 № 4 (92) октябрь–декабрь 2016 г.

Таблица 2. *Результаты финального распознавания полей паспорта РФ для исследованных методов сегментации*

Алгоритм сегментации	Доля полностью верно распознанных полей , %		
	Фамилия	Имя	Отчество
Классический	85.69	90.69	90.15
Обучаемый, сверточная сеть	68.53	76.00	78.30
Обучаемый, рекуррентная сеть на выходах сверточной сети	86.23	90.69	91.38

Таблица 3. *Результаты по дополнительным метрикам качества сегментации*

Алгоритм сегментации	Метрики качества сегментации , %			
	Среднее расстояние	Точность	Полнота	*F1*-мера
Классический	3.97	95.87	95.72	95.78
Обучаемый, сверточная сеть	4.47	95.65	96.89	96.16
Обучаемый, рекуррентная сеть на выходах сверточной сети	4.07	96.75	96.76	96.75

того, нейронная сеть, предназначенная для распознавания полей, была обучена на символах, подвергнутых обжатию, а разрезы, выставленные обучаемым алгоритмом сегментации, передавались без дополнительной обработки и обжатия в систему распознавания. *Таблица 2* содержит экспериментальные результаты точности финального распознавания по полям (доля полностью верно распознанных полей).

Из *таблицы 2* видно, что добавление рекуррентной сети на выходах сверточной сети в подсистеме сегментации сильно повышает точность распознавания полей. Отметим, что в процесс обработки входит нахождение границ паспорта в естественных условиях при съемке с мобильных устройств, чем объясняется неидеальная точность распознавания. Поэтому была выбрана наиболее неблагоприятная выборка с точки зрения возникающих при съемке искажений и т.п.

Таблица 3 содержит экспериментальные результаты методов сегментации на дополнительных метриках, не зависящих от финального классификатора системы распознавания.

Под метрикой среднего расстояния понимается среднее расстояние от найденных разрезов в строке до ближайшего идеального разреза, нормированное на среднюю ширину между идеальными разрезами в этой же строке. Отметим, что обучаемые методы проводят масштабирование изображения к фиксированной ширине, из-за чего показатель среднего расстояния становится хуже даже при идеальной работе сегментатора с масштабированным изображением.

В экспериментах с обучаемыми методами сегментации использовалась реализация нейронных сетей из пакета Lasagne на языке Python, обучение проводилось с помощью алгоритма ADAGRAD [22], также использовалась ранняя остановка при продолжительном отсутствии уменьшения ошибки на валидационной части выборки.

Заключение

В работе были рассмотрены два метода сегментации печатных текстовых полей и проведен их сравнительный анализ на примере модуля сегментации системы распознавания паспорта гражданина Российской Федерации. Первый метод сочетает клас-

сические «инженерные» подходы к сегментации и требует большого количества ручной настройки под конкретные типы документов. Второй метод использует подходы машинного обучения (искусственные нейронные сети) практически на всех стадиях работы, что делает процесс настройки под новые типы полей и документов полностью автоматическим при условии наличия обучающей выборки, чем обусловливается перспективность метода.

В качестве дальнейших исследований метода сегментации, основанного на подходах машинного обучения, планируется проведение экспериментов на других типах полей и документов, анализ и классификация ошибок с целью формирования новых способов расширения обучающей выборки, а также профилирование и оптимизация производительности метода на мобильных устройствах, например путем снижения количества обучаемых параметров. Также интерес представляет исследование методов цельного распознавания целых слов или полей без сегментации с помощью рекуррентных нейронных сетей.

Литература

1. ***R.G. Casey, E. Lecolinet***
   IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, **18**(7), 690. DOI: 10.1109/34.506792.
2. ***C. Patel, A. Patel, D. Shah***
   International Journal of Current Engineering and Technology, 2013, **3**(5), 2075.
3. ***A. Kumar, M. Yadav, T. Patnaik, B. Kumar***
   IJEAT, 2013, **2**(3), 569.
4. ***C. Mancas-Thillou, B. Gosselin***
   В *Proc. 18th International Conference on Pattern Recognition ICPR'06, 4-Vol. Ed., (Hong Kong, 20–24 August, 2006),* Vol. 2, IEEE Publ., 2006, pp. 901–904. DOI: 10.1109/ICPR.2006.362.
5. ***M. Agrawal, D. Doermann***
   В *Proc. The Eighth IAPR International Workshop on Document Analysis Systems DAS'08 (Japan, Nara, 16–19 September, 2008),* IEEE Publ., 2008, pp. 183–190. DOI: 10.1109/DAS.2008.67.
6. ***Y. Lu***
   Pattern Recognition, 1995, **28**(1), 67. DOI: 10.1016/0031-3203(94)00068-W.
7. ***J. Wang, J. Jean***
   В *Proc. 1993 ACM/SIGAPP Symposium on Applied Computing: States of the Art and Practice SAC'93 (USA, IN, Indianapolis, 14–16 February, 1993),* USA, NY, New York, ACM Publ., 1993, pp. 762–769. DOI: 10.1145/162754.168698.
8. ***A. Graves, S. Fernández, F. Gomez, J. Schmidhuber***
   В *Proc. 23rd International Conference on Machine learning ICML'06 (USA, PA, Pittsburgh, 25–29 June, 2006),* USA, NY, New York, ACM Publ., 2006, pp. 369–376. DOI: 10.1145/1143844.1143891.
9. ***T.M. Breuel, A. Ul-Hasan, M.A. Al-Azawi, F. Shafait***
   В *Proc. 12th International Conference on Document Analysis and Recognition (ICDAR), (USA, DC, Washington, 25–28 August, 2013),* IEEE Publ., 2013, pp. 683–687. DOI: 10.1109/ICDAR.2013.140.
10. ***S. Hochreiter, J. Schmidhuber***
    Neural Computation, 1997, **9**(8), 1735. DOI: 10.1162/neco.1997.9.8.1735.
11. ***C.J.C. Burges, J.I. Ben, C.R. Nohl***
    В *«The Postal Service in the Information Age» Proc. USPC 5th Advanced Technology Conference, (USA, DC, Washington, 30 November – 2 December, 1992)*, 1992, p. A-117–A-124.
12. ***J. Duchi, E. Hazan, Y. Singer***
    JMLR, 2011, **12**, 2121.
13. ***A. Krizhevsky, I. Sutskever, G.E. Hinton***
    В *Advances in Neural Information Processing Systems 25 (Proc. NIPS 2012), (USA, NV, Stateline, 3–8 December, 2012),* USA, NY, Red Hook, Publ. Curran Associates, Inc., 2012, pp. 1097–1105.
14. ***P. dos Santos, G.S. Clemente, T.I. Ren, G.D. Cavalcanti***
    В *Proc. 10th International Conference on Document Analysis and Recognition (ICDAR), (Spain, Catalonia, Barcelona, 26–29 July, 2009),* IEEE Publ., 2009, pp. 651–655. DOI: 10.1109/ICDAR.2009.183.
15. ***В.Л. Арлазаров, П.А. Куратов, О.А. Славин***
    В *Методы и средства работы с документами, сер. Труды ИСА РАН, под ред. В.Л. Арлазарова, Н.Е. Емельянова,* Москва**,** УРСС, 2000, с. 31–51.
16. ***В.Л. Арлазаров, П.А. Куратов, О.А. Славин***
    В *Организационное управление и искусственный интеллект, сер. Труды ИСА РАН, под ред. В.Л. Арлазарова,* Москва, УРСС, 2003, с. 176–184.
17. ***В.В. Арлазаров, А.Е. Жуковский, В.Е. Кривцов, Д.П. Николаев, Д.В. Полевой***
    Информационные технологии и вычислительные системы, 2014, №3, 71.
18. ***А. Иванова, Е. Кузнецова, Д. Николаев***
    В *Сб. труд. 39-й Междисципл. шк.-конф. ИТиС 2015 «Информационные технологии и системы 2015», (РФ, Сочи, 7–11 сентября, 2015 г.),* Москва, Изд. ИППИ им. А.А. Харкевича РАН, 2015, с. 1169–1184.
19. ***В.И. Левенштейн***
    Проблемы передачи информации, 1965, **1**(1), 8.
20. ***M.-C. Jung, Y.-C. Shin, S.N. Srihari***
    В *Proc. IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC'99), (Japan, Tokyo, 12–15 October, 1999),* IEEE Publ., 1999, Vol. VI, pp. 863–867. DOI: 10.1109/ICSMC.1999.816665.
21. ***J.H. Bae, K.C. Jung, J.W. Kim, H.J. Kim***
    Pattern Recognition Letters, 1998, **19**(8), 701. DOI: 10.1016/S0167-8655(98)00048-8.
22. ***D. Wen, X. Ding***
    В *Proc. SPIE 5296, Document Recognition and Retrieval XI, (USA, CA, San Jose, 21–22 January, 2003),* SPIE Press, 2003, pp. 147–154. DOI: 10.1117/12.528951.

English

Research of Segmentation Methods for Images of Document Textual Blocks Based on the Structural Analysis and Machine Learning*

Timofey S. Chernov –
National University of Science and Technology MISIS
4, Leninskiy Ave., Moscow, 119049, Russia
e-mail: chernov.tim@gmail.com

Dmitriy A. Ilin –
Institute for System Analysis FRC
"Computer Science and Control" RAS
9, 60-letiya Octyabrya Ave., Moscow, 117312, Russia
e-mail: dmitry.ilin@phystech.edu

Pavel V. Bezmaternykh –
Smart Engines Service Ltd.
9, 60-letiya Octyabrya Ave.,
Moscow, 117312, Russia
e-mail: bezmpavel@smartengines.biz

Igor A. Faradzhev –
Institute for System Analysis FRC
"Computer Science and Control" RAS
9, 60-letiya Octyabrya Ave.,
Moscow, 117312, Russia
e-mail: ifardjev@yahoo.com

Simon M. Karpenko –
A.A. Kharkevich Institute for Information
Transmission Problems RAS
19-1, Bolshoy Karetny Per.,
Moscow, 127051, Russia
e-mail: simon.karpenko@gmail.com

Abstract

Text image segmentation methods are the most important components of text and document optical recognition systems. In this paper two methods of printed text fields' segmentation are discussed. The first considered method develops classical approaches to text segmentation and comprises such stages as projection analysis, preliminary cutting and dynamic programming taking into account the probability scores of character recognition. The second method uses extensively the machine learning approaches, particularly convolutional and recurrent neural nets, that allows developing the segmentation algorithms without numerous heuristics methods tied to specific document field types and also increases these algorithms robustness to various distortions occurring while videoshooting by means of mobile devices. The authors performed comparative analysis of these two methods through the example of segmentation module included in the recognition system of Russian Federation citizen's passport.

Keywords: optical character recognition, document analysis, segmentation, machine learning, videostream analysis.

Images & Tables

Fig. 1. Images of the Russian Federation citizen's passport fields with different distortions. These examples show difficulties of segmentation task applied to images captured by mobile cam in poor conditions.

* *The work was financially supported by RFBR (projects 13-07-12172, 13-07-12173 and 14-07-00730).*

Fig. 2. An area cutted from the passport field.

Fig. 3. Filtered area of the passport field.

Fig. 4. The area correction based on the support lines.

Fig. 5. An image projection.

Fig. 6. The primary arrangement of cross-sections.

Fig. 7. The result of letters compression and noise removal.

Fig. 8. The result of the subsystem operation.

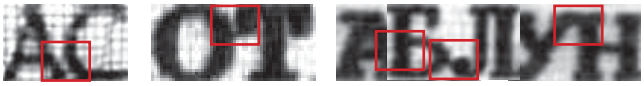


Fig. 9. Examples of the difficulties while symbols recognition.

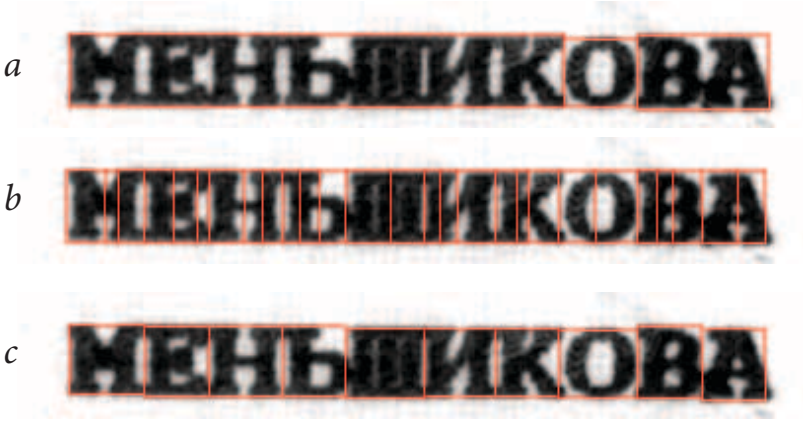


Fig. 10. The example of the second work stage of the algorithm: a – initial cross-sections; b – the result of the two groups of cross-sections combination; c – the final result of segmentation.

DOI: 10.22204/2410-4639-2016-092-04-55-71 № 4 (92) октябрь–декабрь 2016 г.

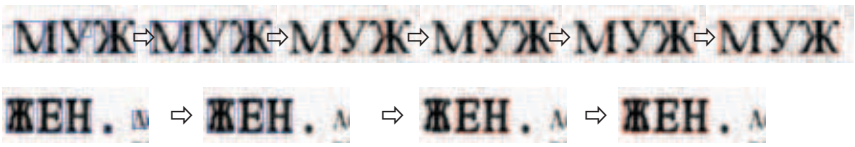


Fig. 11. Stages of the "Gender" field processing.

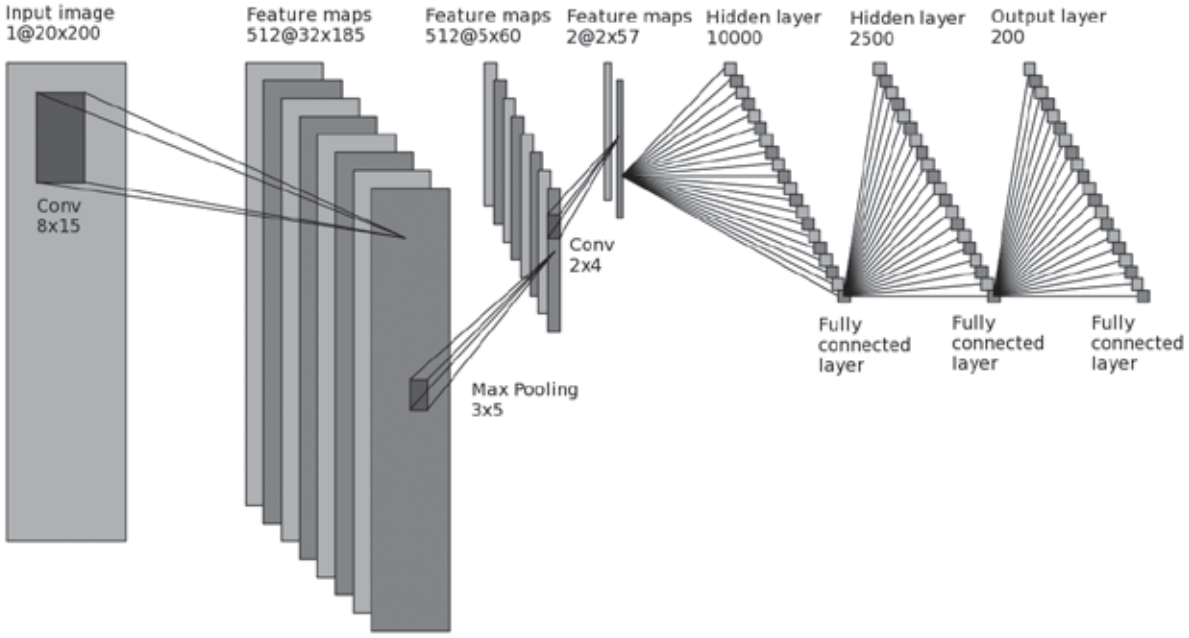


Fig. 12. The convolutional neural network structure of the algorithm first stage. The outputs of this stage are the probabilistic score of cross-sections for different parts of image.

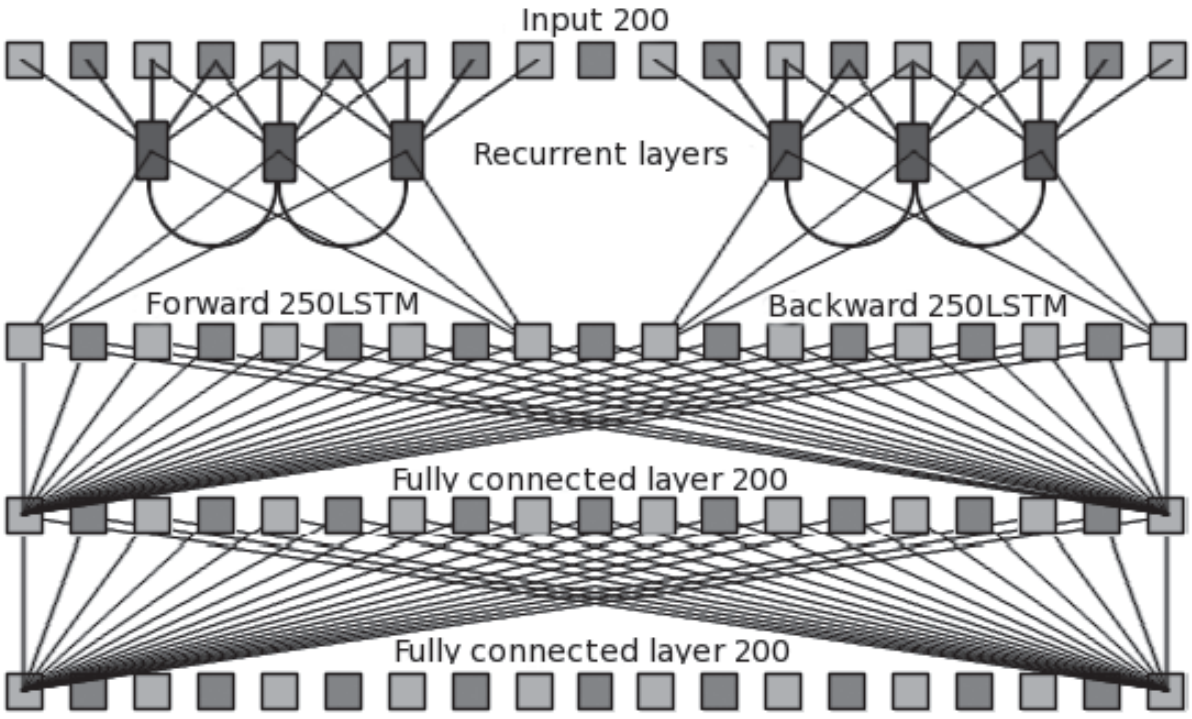
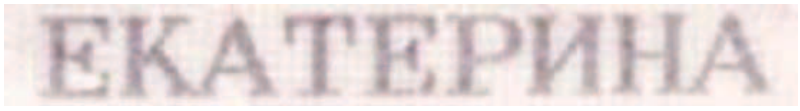
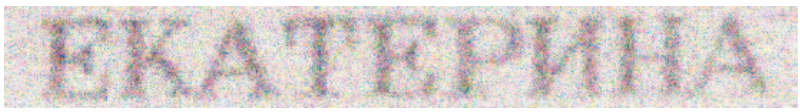
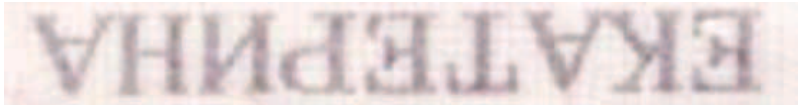
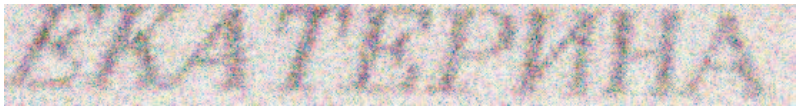


Рис. 13. The recurrent neural network structure of the algorithm second stage, inputs of which are the outputs (probabilistic scores) of the convolutional neural network.

Table 1. *Conversion of the images for the augmentation of learning sampling*

Conversion	Illustration
Original image	
Gaussian noise	
Projective distortion	
Gaussian blur	
Shifts	
Mixing letters	
Reflections	
Wricks	
Distortions combinations	

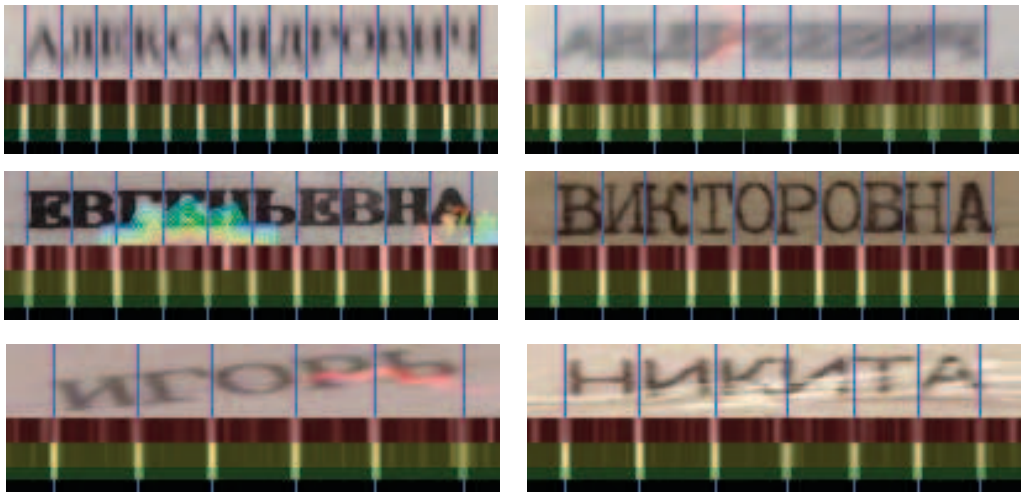


Fig. 14. *The work example of the segmentation learning method with the output of intermediate results.*

DOI: 10.22204/2410-4639-2016-092-04-55-71 № 4 (92) октябрь–декабрь 2016 г.

Table 2. *The final results of fields recognition for different segmentation algorithms*

Segmentation algorithm	Proportion of entirely correctly recognized fields, %		
	Surname	Name	Patronym
Classic (dynamic programming)	85.69	90.69	90.15
Convolutional neural network	68.53	76.00	78.30
Convolutional + reccurent neural network	86.23	90.69	91.38

Table 3. *Precision score of the segmentation algorithms in accordance with different metrics of segmentation quality*

Segmentation algorithm	Metrics of segmentation quality, %			
	Mean distance	Precision	Recall	*F1*-score
Classic (dynamic programming)	3.97	95.87	95.72	95.78
Convolutional neural network	4.47	95.65	96.89	96.16
Convolutional + reccurent neural network	4.07	96.75	96.76	96.75

References

1. ***R.G. Casey, E. Lecolinet***
IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, **18**(7), 690. DOI: 10.1109/34.506792.
2. ***C. Patel, A. Patel, D. Shah***
International Journal of Current Engineering and Technology, 2013, **3**(5), 2075.
3. ***A. Kumar, M. Yadav, T. Patnaik, B. Kumar***
IJEAT, 2013, **2**(3), 569.
4. ***C. Mancas-Thillou, B. Gosselin***
In *Proc. 18th International Conference on Pattern Recognition ICPR'06, 4-Vol. Ed., (Hong Kong, 20–24 August, 2006),* Vol. 2, IEEE Publ., 2006, pp. 901–904. DOI: 10.1109/ICPR.2006.362.
5. ***M. Agrawal, D. Doermann***
In *Proc. The Eighth IAPR International Workshop on Document Analysis Systems DAS'08 (Japan, Nara, 16–19 September, 2008),* IEEE Publ., 2008, pp. 183–190. DOI: 10.1109/DAS.2008.67.
6. ***Y. Lu***
Pattern Recognition, 1995, **28**(1), 67.
DOI: 10.1016/0031-3203(94)00068-W.
7. ***J. Wang, J. Jean***
In *Proc. 1993 ACM/SIGAPP Symposium on Applied Computing: States of the Art and Practice SAC'93 (USA, IN, Indianapolis, 14–16 February, 1993),* USA, NY, New York, ACM Publ., 1993, pp. 762–769. DOI: 10.1145/162754.168698.
8. ***A. Graves, S. Fernández, F. Gomez, J. Schmidhuber***
In *Proc. 23rd International Conference on Machine learning ICML'06 (USA, PA, Pittsburgh, 25–29 June, 2006),* USA, NY, New York, ACM Publ., 2006, pp. 369–376. DOI: 10.1145/1143844.1143891.
9. ***T.M. Breuel, A. Ul-Hasan, M.A. Al-Azawi, F. Shafait***
In *Proc. 12th International Conference on Document Analysis and Recognition (ICDAR), (USA, DC, Washington, 25–28 August, 2013),* IEEE Publ., 2013, pp. 683–687. DOI: 10.1109/ICDAR.2013.140.
10. ***S. Hochreiter, J. Schmidhuber***
Neural Computation, 1997, **9**(8), 1735.
DOI: 10.1162/neco.1997.9.8.1735.
11. ***C.J.C. Burges, J.I. Ben, C.R. Nohl***
In *«The Postal Service in the Information Age» Proc. USPC 5th Advanced Technology Conference, (USA, DC, Washington, 30 November – 2 December, 1992)*, 1992, p. A-117–A-124.
12. ***J. Duchi, E. Hazan, Y. Singer***
JMLR, 2011, **12**, 2121.
13. ***A. Krizhevsky, I. Sutskever, G.E. Hinton***
In *Advances in Neural Information Processing Systems 25 (Proc. NIPS 2012), (USA, NV, Stateline, 3–8 December, 2012),* USA, NY, Red Hook, Publ. Curran Associates, Inc., 2012, pp. 1097–1105.
14. ***P. dos Santos, G.S. Clemente, T.I. Ren, G.D. Cavalcanti***
In *Proc. 10th International Conference on Document Analysis and Recognition (ICDAR), (Spain, Catalonia, Barcelona, 26–29 July, 2009),* IEEE Publ., 2009, pp. 651–655. DOI: 10.1109/ICDAR.2009.183.
15. ***V.L Arlazarov, P.A Kuratov, O.A. Slavin***
In *Methods and Means of Work with Documnts. Ser. Proc. of Institute for System Analysis RAS [Metody i sredstva raboty s documentami. Ser. Trudy ISA RAN],* Eds V.L. Arlazarov, N.E. Emelyanov, Moscow, URSS Publ., 2000, pp. 31–51 (in Russian).
16. ***V.L Arlazarov, P.A Kuratov, O.A. Slavin***
In *Organizational Management and Artificial Intelligence. Ser. Proc. of Institute for System Analysis RAS [Organizatsionnoe upravlenie i iskusstvennyy intellekt. Ser. Trudy ISA RAN],* Ed. V.L. Arlazarov, Moscow, URSS Publ., 2003, pp. 176–184 (in Russian).
17. ***V.L Arlazarov, A.E. Zhukovsky, V.E. Krivtsov, D.P. Nikolaev, D.V. Polevoy***
J. Information Technology and Computer Systems [Informatsionnye tekhnologii i kompyuternye sistemy], 2014, №3, 71 (in Russian).
18. ***A.Ivanova, E. Kuznetsova, D. Nikolaev***
In *Proc. 39th School-Conf. IT&S 2015 "Information Technologies and Systems 2015" [Informatsionnye tekhnologii i sistemy], (RF, Sochi, 7–11 September, 2015),* RF, Moscow, A.A. Kharkevich IITP RAS Publ., 2015, pp. 1169–1184 (in Russian).
19. ***V.I. Levenshtein***
Problems of Information Tansmission: A Translation of Problemy Peredachi Informatsii, 1965, **1**(1), 8.
20. ***M.-C. Jung, Y.-C. Shin, S.N. Srihari***
In *Proc. IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC'99), (Japan, Tokyo, 12–15 October, 1999),* IEEE Publ., 1999, Vol. VI, pp. 863–867. DOI: 10.1109/ICSMC.1999.816665.
21. ***J.H. Bae, K.C. Jung, J.W. Kim, H.J. Kim***
Pattern Recognition Letters, 1998, **19**(8), 701.
DOI: 10.1016/S0167-8655(98)00048-8.
22. ***D. Wen, X. Ding***
In *Proc. SPIE 5296, Document Recognition and Retrieval XI, (USA, CA, San Jose, 21–22 January, 2003),* SPIE Press, 2003, pp. 147–154.
DOI: 10.1117/12.528951.

Формирование ошибки в методе компьютерной томографии: от проекции до интерпретации результата*

М.В. Чукалина, Д.П. Николаев, А.В. Бузмаков, А.С. Ингачева, Д.А. Золотов, А.П. Гладков, В.Е. Прун, Б.С. Рощин, И.А. Щелоков, В.И. Гулимова, С.В. Савельев, В.Е. Асадчиков

В работе впервые сделана попытка описать весь спектр источников ошибок, искажающих томографическое изображение в процессе формирования, провести классификацию источников и установить качественную связь между причинами искажений. Применение метода компьютерной томографии в таких областях, как медицина и промышленная диагностика делает цену ошибки при интерпретации результата непомерно высокой. Поэтому конечная цель начатой работы заключается в построении математического выражения, которое связывает величину финальной ошибки с параметрами каждого из формирующих ее этапов. Все представленные в статье томографические изображения получены авторами и являются результатами экспериментов, проведенных на аппаратно-программных комплексах, функционирующих в институте кристаллографии им. А.В. Шубникова РАН.

Ключевые слова: метод компьютерной томографии, формирование изображения, источники финальной ошибки, взаимосвязь источников.

** Работа выполнена при финансовой поддержке РФФИ (проекты №№ 13-07-00970, 13-07-12179, 13-02-00552 и 16-04-00815).*

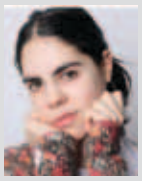
Введение

Еще одно-два десятилетия назад при слове «компьютерный томограф» лишь на лице исследователя можно было прочесть интерес. Сегодня словом «томограф» уже мало кого удивишь. Компьютерные томографы применяются в медицине – как для доклинической визуализации, так и в целях наблюдения за динамикой результатов врачебного вмешательства, в промышленности – для контроля качества

ЧУКАЛИНА
Марина Валерьевна
ФНИЦ «Кристаллография и фотоника» РАН

НИКОЛАЕВ
Дмитрий Петрович
Институт проблем передачи информации им. А.А. Харкевича РАН

БУЗМАКОВ
Алексей Владимирович
ФНИЦ «Кристаллография и фотоника» РАН



ИНГАЧЕВА
Анастасия Сергеевна
ФНИЦ «Кристаллография и фотоника» РАН, НИУ «Высшая школа экономики»

ЗОЛОТОВ
Денис Александрович
ФНИЦ «Кристаллография и фотоника» РАН

ГЛАДКОВ
Андрей Павлович
Институт проблем передачи информации им. А.А. Харкевича РАН

ПРУН
Виктор Евгеньевич
Московский физико-технический институт (государственный университет)

РОЩИН
Борис Сергеевич
ФНИЦ «Кристаллография и фотоника» РАН

ЩЕЛОКОВ
Игорь Александрович
Институт проблем технологии микроэлектроники и особочистых материалов РАН

ГУЛИМОВА
Виктория Игоревна
Научно-исследовательский институт морфологии человека

САВЕЛЬЕВ
Сергей Вячеславович
профессор, Научно-исследовательский институт морфологии человека

АСАДЧИКОВ
Виктор Евгеньевич
профессор, ФНИЦ «Кристаллография и фотоника» РАН

на разных этапах технологических процессов, и по-прежнему остаются важным инструментом для проведения научных исследований, особенно при изучении невидимого для человека наномира. За несколько десятилетий мы научились использовать томографы для наблюдения за динамическими процессами, протекающими в реальном времени и в некоторых случаях даже управлять этими процессами. Для чего же тогда были проведены исследования, результаты которых представлены в статье? Ответ достаточно прост: чтобы сокращалось число ложных врачебных результатов, реже случались сбои в космическом (и не только) приборостроении и еще, чтобы приоткрыл свои двери для нас мир невидимого человеческим глазом. Томограф – прибор, который наряду с измерительной частью уже наделен и встроенной системой обработки результатов проводимых измерений. Данная система в обязательном порядке включает этап реконструкции, т.е. расчет пространственного распределения изучаемой характеристики или характеристик измеряемого объекта. Допустим, мы уже знакомы с объектом предполагаемого исследования, мы его видели снаружи и знаем его форму. С помощью томографа мы заглядываем внутрь, не разрушая объект. Мы поймем всю мощь метода томографии, если ответим себе на вопрос: что мы хотим найти, заглядывая внутрь? Для промышленного контроля ответ будет одним (трещина, нарушение размеров и пр.), для медицины ответ будет лежать в другой плоскости (размер новообразования, динамика процесса вживления импланта, гемодинамические нарушения и пр.). Задача ученого – количественно описать интересующую характеристику объекта так, чтобы, с одной стороны, удалось построить математическое выражение, связывающее результат измерения с данной характеристикой, а с другой – так представить полученное распределение, чтобы пользователь томографа сумел конструктивно применить результат компьютерной томографии; врач смог назначить лечение, биолог – построить модель динамики развития наблюдаемой живой системы, инженер – поменять параметры технологического процесса и пр. Однако цена ошибки принятого специалистом решения в ряде случаев оказывается достаточно высокой, поэтому, принимая решение, специалист должен иметь информацию о достоверности используемых томографических изображений. Достоверность характеризуется вероятностью того, что истинное значение величины единичного элемента 2D/3D-изображения находится в указанных пределах. Методов расчета данной вероятности в мире пока не предложено. В данной статье мы описываем весь спектр источников ошибок, искажающих томографическое изображение в процессе формирования, классифицируем источники и в заключение устанавливаем качественную связь между причинами искажений. Все представленные в статье томографические изображения получены авторами и являются результатами экспериментов, проведенных на прототипах томографических комплексов, созданных и функционирующих в Институте кристаллографии им. А.В. Шубникова ФНИЦ «Кристаллография и фотоника» РАН (ИК РАН) [1] *(рис. 1).*

Томографический эксперимент и погрешности измерений

В ходе выполнения проектов РФФИ №№ 13-07-00970 и 13-07-12179 мы вели активные работы по развитию программной части комплекса, которая сегодня

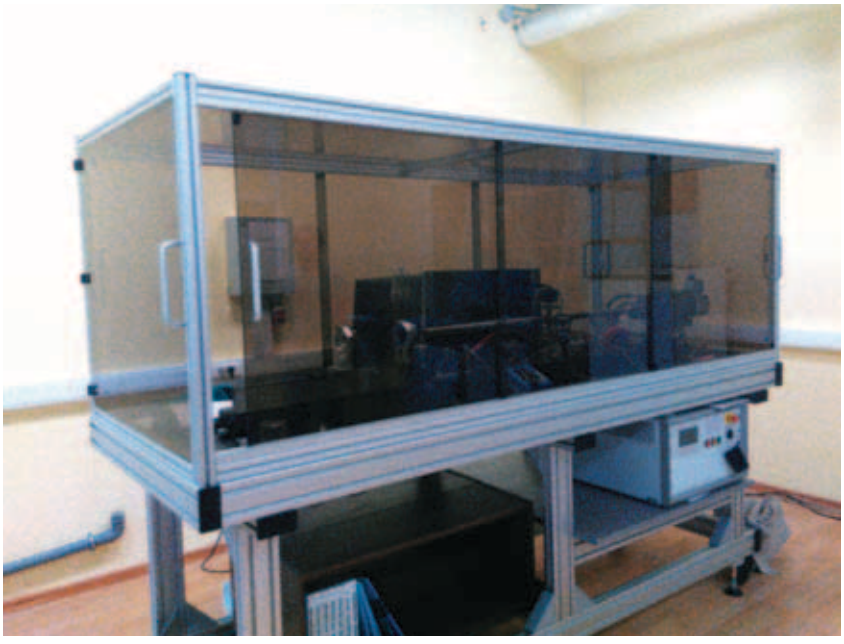


Рис. 1. *Аппаратно-программный томографический комплекс ИК РАН.*

включает модули как удаленного управления и контроля за экспериментами, так и для работы с результатами измерений.

Постоянно идет развитие и аппаратной части комплекса. Оно сопровождается решением возникающих задач, связанных с математическим моделированием [2]. Введение дополнительных оптических элементов в схему позволило реализовать измерения

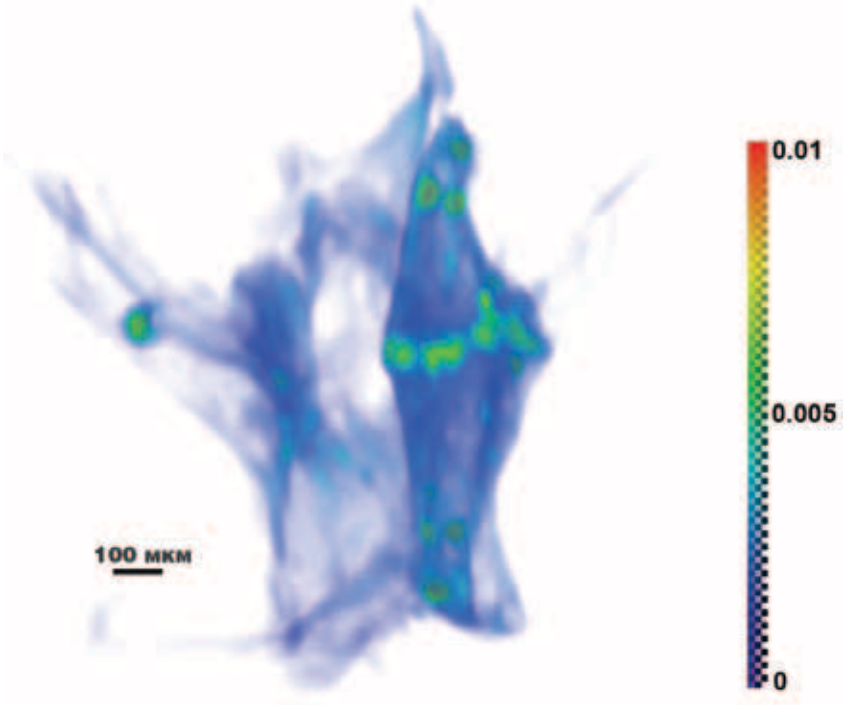


Рис. 2. *Результат томографической реконструкции позвонка геккона.*

Рис. 3. *Фотография природного алмаза.*

с увеличением. Достигнуто микронное пространственное разрешение. Непосредственный доступ к оборудованию и «сырым» результатам измерений, проводимых в требуемых режимах, позволяет нам осуществлять как плановые, так и тестовые измерения. Последние необходимы нам, чтобы оценить погрешность проводимых измерений E_m^d. Данным обозначением мы вводим первый источник ошибок, связав его с используемым в эксперименте оборудованием.

Для демонстрации проводимых плановых измерений на *рисунке 2* представлен результат реконструкции проксимального хвостового позвонка хрящепалого геккона. Исследования, проводимые совместно с группой ученых из Научно-исследовательского института морфологии человека, касались проблемы деминерализации костной ткани в условиях невесомости [3].

Ранее нами было показано [4], что метод рентгеновской дифракционной томографии (топо-томографии), предложенный для исследования слабопоглощающих кристаллов, например природного алмаза *(рис. 3)*, может быть реализован не только на синхротроне, но и с использованием лабораторных источников рентгеновского излучения.

В рамках проекта РФФИ №13-02-00552 были выполнены работы по развитию данного метода. Прототип топо-томографа, также созданный и функционирующий в ИК РАН, позволяет одновременно проводить измерение ослабленного на объекте зондирующего и дифрагированного пучков *(рис. 4)*. На *рисунке 5* представлены результаты компьютерной томографии (пространственное распределение линейного коэффициента ослабления рентгеновского излучения природным алмазом, *рис. 5a*) и результат топо-томографии – пространственное распределение дислокаций *(рис. 5b)* [5]. В методе топо-томографии выбор направления оси

вращения согласован с направлением вектора обратной решетки одной из отражающих плоскостей кристалла. Итак, помимо оборудования, вовлеченного в процесс формирования томографических проекций, возникает еще один класс ошибок. Данный класс связан с геометрией организации сбора проекций. В зависимости от схемы формирования зондирующего пучка она может быть параллельной или конусной.

Исследуемый объект может оставаться неподвижным, тогда вокруг оси вращения по заданной траектории вращается система «излучатель–детектор» или же система «излучатель–детектор» остается неподвижной, но вращается исследуемый образец. То есть ошибка, назовем ее E_m^g, связана с отклонением идеальных параметров геометрии эксперимента (положение и наклон оси вращения относительно направления распространения зонда, расстояние «излучатель–образец», расстояние «образец–детектор», стабильность положения излучателя на оси, проходящей через условные центры объекта и детектора) от параметров, используемых при решении задачи реконструкции. И мы готовы перейти от этапа измерения к этапу реконструкции. Однако нам не удастся этого сделать без введения

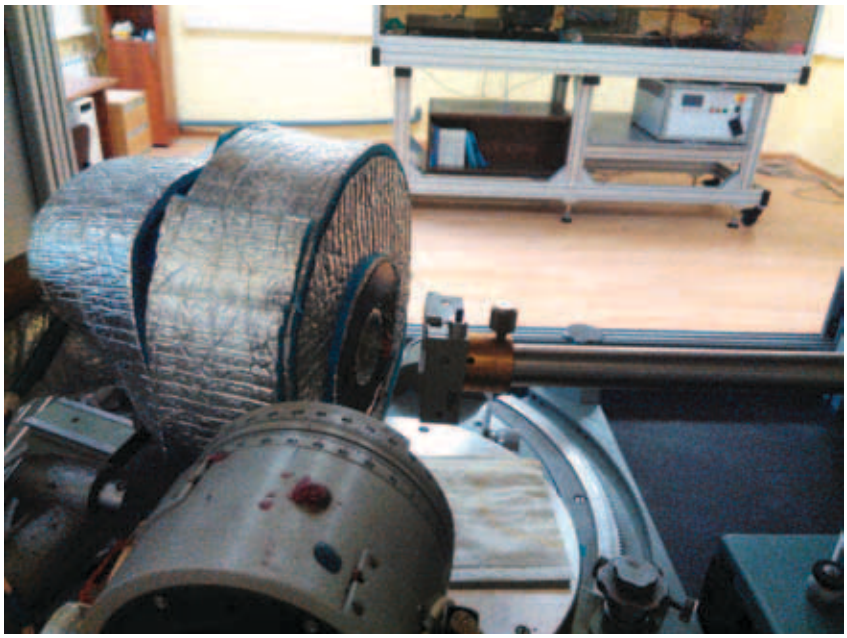


Рис. 4. Измерительный блок для топо-томографических измерений.

математической модели формирования томографической проекции.

Задача реконструкции

В методе компьютерной томографии объект зондируется рентгеновским излучением. Закон Бугера–Ламберта–Бера описывает процесс ослабления монохроматического рентгеновского излучения, проходящего через однородный слой материала толщины *l*:

$$I(l) = I_0 exp(-\mu l). \quad (1)$$

Здесь $I(l)$ – интенсивность прошедшего через объект монохроматического излучения (фотонов в секун-

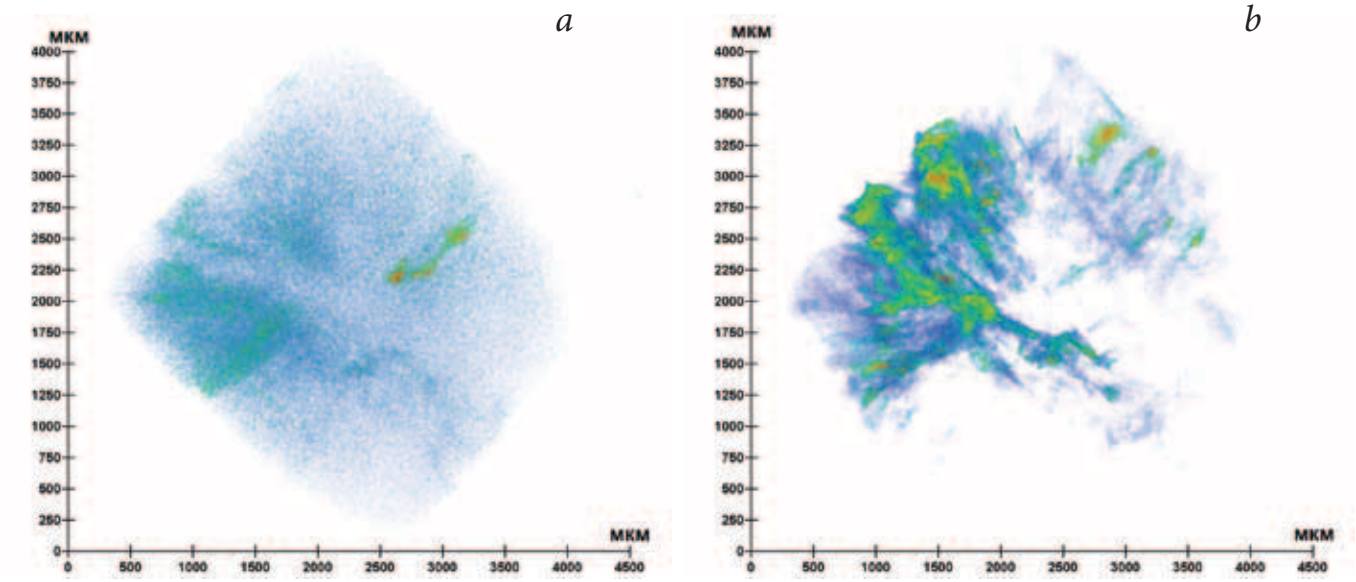


Рис. 5. Результаты компьютерной томографии алмаза, изображенного на рисунке 3 (a) и топо-томографии того же алмаза (b).

ду), I_0 – начальная интенсивность рентгеновского излучения, μ – линейный коэффициент ослабления. Считаем пучок бесконечно тонким. Направление зондирования перпендикулярно поверхности слоя. Легко показать, что если слой бесконечно тонкий, то μ есть относительное ослабление рентгеновского излучения на единице длины слоя однородного материала. То есть результатом применения метода компьютерной томографии является пространственное распределение линейных коэффициентов ослабления в объеме исследуемого объекта. Связь линейных коэффициентов ослабления с характеристиками, которыми мы хотели бы описать, например параметрами человеческого тела, не вполне очевидна. Для визуальной оценки плотности анатомических структур методом компьютерной томографии используется шкала ослабления рентгеновского излучения, названная шкалой Хаунсфилда, где полученное при данном спектре падающего излучения значение поглощения исследуемой ткани (μ), соотносится с поглощением воды при тех же условиях (μ_{water}). Эта шкала названа по имени английского инженера, получившего вместе с теоретиком Алланом Кормаком в 1979 г. Нобелевскую премию по физиологии и медицине «за разработку компьютерной томографии». Шкала представляет собой линейное преобразование

$$HU = 1000 \times \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}}. \qquad (2)$$

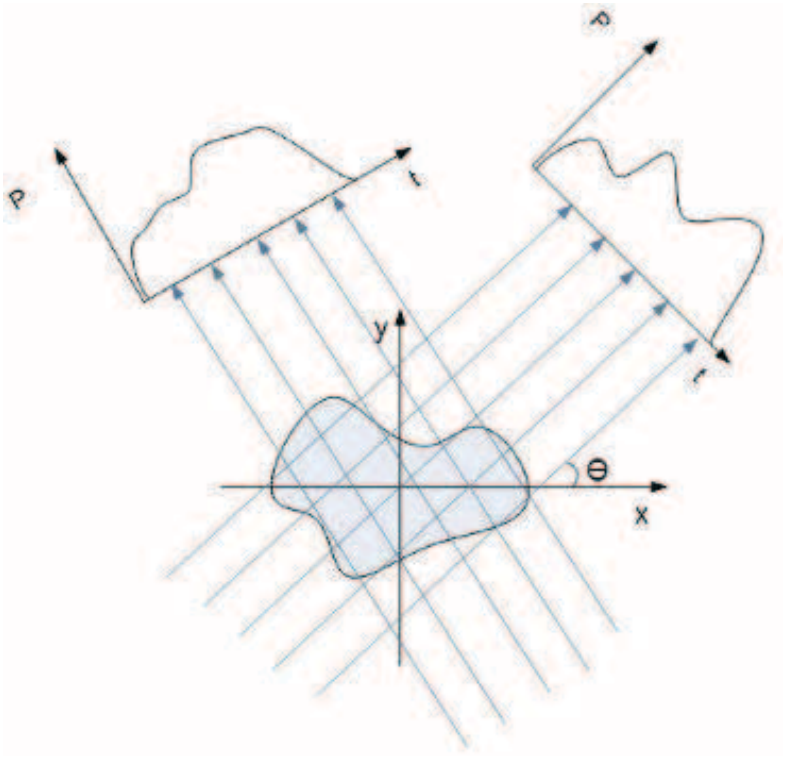


Рис. 6. Параллельная схема измерений.

Получаемые с медицинских компьютерных томографов изображения (2D-срезы или 3D-объемы) представляют собой распределение величин Хаунсфилда (так называемых КТ-чисел) в формате DICOM (Digital Imaging and Communications in Medicine) [6]. Приведем в качестве примера диапазоны КТ-чисел для разных тканей человеческого тела, представленные в медицинской энциклопедии. Головной мозг – от +2 до +30 HU, кости – от +150 до +1000 HU, легкие – от -400 до -700 HU [7], т.е. переход к КТ-числам и есть та самая модель, в рамках которой описываются анатомические объекты.

Чтобы записать математическое выражение, которое связывает величину ослабленной на пути зондирования интенсивности рентгеновского излучения с распределением линейного коэффициента ослабления, обратимся к схеме томографического эксперимента. Для простоты рассмотрим параллельный пучок и бесконечнотонкое сечение 2D-объекта *(рис. 6)*. Центр Декартовой системы координат совпадает с центром вращения. Стрелками показаны направления зондирования. Объект выделен голубым цветом.

Бесконечно тонкий рентгеновский луч $I_0(t,\theta)$ проходит сквозь объект $\mu(x,y)$ под углом θ на расстоянии t от центра вращения и регистрируется позиционно-чувствительным детектором $I(t,\theta)$:

$$I(t,\theta) = I_0(t,\theta) exp(-\iint \mu(x,y) \cdot \cdot \delta(x\cos\theta + y\sin\theta - t)dxdy). \qquad (3)$$

Назовем пространство I пространством измерений. Сделав предположение, что $I_0(t,\theta) = const$, и прологарифмировав отношение, придем к выражению для точки томографической проекции:

$$P(t,\theta) = ln(\frac{I_0}{I(t,\theta)}) = -\iint \mu(x,y) \cdot \cdot \delta(x\cos\theta + y\sin\theta - t)dxdy). \qquad (4)$$

То есть томографическая проекция 2D – вектор значений P при фиксированном проекционном угле θ. В непрерывном случае преобразо-

вание Радона [8], которое переводит функцию во множество ее интегралов по направлениям, а именно оно записано в выражении (4), обратимо. Если пространство Радона P полное, то существует преобразование, которое позволяет однозначно восстановить функцию μ, если известна функция I. Поскольку проекции измеряются лишь по некоторым выбранным направлениям, чтобы минимизировать величину поглощенной объектом дозы, т.е. пространство Радона не является полным, величина ошибки, связанной с выбранным методом реконструкции E_r^{tech}, возникают именно на этом месте. Существует несколько принципиально разных подходов к выбору метода реконструкции [9, 10]. Основное различие интегральных и алгебраических методов реконструкции состоит в том, что алгоритмы, основанные на интегральном подходе, считают восстанавливаемую функцию непрерывной вплоть до этапа вычислений, а алгебраические методы работают с дискретным изображением (кусочно-постоянной функцией) уже на этапе постановки задачи. В данной работе мы анализируем эффективную версию алгебраического метода SART [11], выделяя этапы формирования финальной ошибки реконструкции E_r^{tech}.

Перепишем уравнение (4) в операторном виде:

$$P = W\mu. \quad (5)$$

Здесь P – набор томографических проекций (синограмма), μ – искомое изображение, элемент w_{ij} матрицы W описывает вклад i-го пиксела в лучевую сумму, формируемую j-ым лучом. Лучевая сумма – это сумма значений функции вдоль заданного направления. В зависимости от того, какая модель вклада пиксела используется для решения операторного уравнения (бинарная, модель длины или площадная), ошибка расчета лучевой суммы будет разной. Для быстрого расчета лучевых сумм нами было предложено заменить модель, наилучшим образом описывающую физический луч, на модель дискретных симметричных лучей ступенек. При таком приближении преобразование Радона переходит в преобразование Хафа и асимптотическое время, требуемое для расчета одной итерации, удалось сократить с $0(n^3)$ до $0(n^2 log(n))$ [12]. Это значимое ускорение, поскольку линейные размеры изображений составляют несколько тысяч пикселей, но его цена – дополнительная ошибка при расчете лучевой суммы. Однако в работе [13] было показано, что с увеличением размера изображения, максимальное отклонение соответствующего Хафовского паттерна уменьшается и, соответственно, ошибка, вносимая таким приближением, тоже становится пренебрежимо малой. Рассмотрим задачу реконструкции томографических измерений как задачу минимизации целевой функции невязки:

$$(H\mu - p)^2 \rightarrow min, \quad (6)$$

где H – линейный оператор преобразования Хафа, p - томографическая проекция. Выписанная оптимизационная задача решается итеративным методом, например методом наискорейшего спуска. Выпишем шаг очередной итерации ξ при градиентном спуске:

$$\mu^{\xi} = \mu^{\xi-1} + \alpha H^T(H\mu^{\xi-1} - p), \quad (7)$$

где α – параметр релаксации, управляющий ходом минимизационного процесса [14]. В частности, если выбирать релаксационный параметр, используя «жадную» стратегию, то есть такую, при которой невязка должна быть максимально уменьшена вдоль текущего направления, получится метод наи-

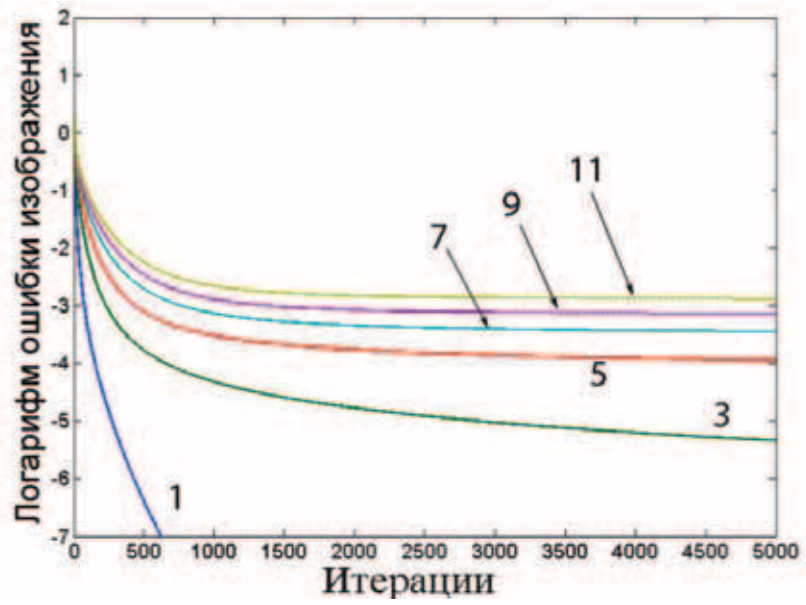


Рис. 7. Поведение среднеквадратичной ошибки реализации алгебраического метода (цифры обозначают степени разреженности).

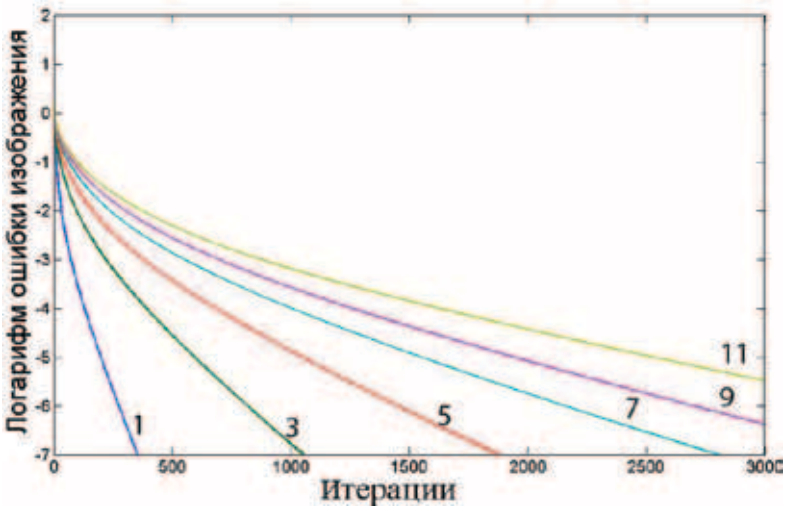


Рис. 8. Поведение среднеквадратичной ошибки реализации алгебраического метода шагом медианной фильтрации (цифры обозначают степени разреженности).

скорейшего спуска. Поведение финальной среднеквадратичной ошибки (6) в зависимости от номера итерации для разных размеров изображения представлено на *рисунке 7*. Реконструкцию выполняли из модельной синограммы, рассчитанной для изображения размером 256×256 пикселов и полного набора проекционных углов (1021) в пространстве Хафа.

На *рисунке 7* представлены кривые как для случая полного пространства Хафа, так и для нескольких случаев разреженных пространств. На рисунке представлены случаи для степеней разреженности 3, 5, 7, 9 и 11. *Рисунок 8* демонстрирует поведение ошибки в условиях присутствия дополнительной регуляризирующей субитерции в итерационной процедуре.

Для выбора или построения оптимальной регуляризирующей процедуры, используемой в субитерации, принимаются во внимание свойства пространства изображения [15]. Так, медианная фильтрация *(рис. 8)* используется для борьбы с шумом, сильно искажающим результаты измерений лабораторных томографов [16, 17].

На этом можно завершить рассмотрение этапа реконструкции и перейти к рассмотрению связи ошибок, возникающих на этапе измерения, и ошибок, формируемых на этапе реконструкции. Однако перед началом отдельно следует рассмотреть случай, когда в синограмме присутствуют пикселы, для которых измеренные значения лежат в районе нуля (в пределах погрешности измерений). Это указание на тот факт, что в объекте существуют сильно поглощающие включения. Встает вопрос о том, как выделить такие области и как оценить ошибку реконструкции в данных областях. Обращение преобразования Радона поместит в соответствующие единичные элементы объема внутри объекта нулевые или случайные значения. Чтобы выделить области исследуемого объема, результаты реконструкции в которых ненадежны, мы предложили рассчитать матрицу Якоби по всему пространству измерений, поскольку, если решение задачи (6) найдено, то из условия

$$\mu_i : \forall_j \frac{\partial I_j}{\partial \mu_j} = 0 \ . \qquad (8)$$

следует, что показание детектора j нечувствительно к изменению состава единичного объема i. Если матрицу Якоби записать в векторной форме, то можно представить ее как изображение, повторяющее изображение объекта, в каждом пикселе которого записан вектор. Координаты последнего – частные производные измерений по всем направлениям зондирования. Выбор метода анализа и визуализации векторного изображения является самостоятельной задачей. В данной работе мы приведем несколько способов визуализации для того, чтобы продемонстрировать необходимость оптимального выбора и создания автоматизированной системы в дальнейшем.

Несколько результатов визуализации векторного изображения представлены на *рисунке 9*. Для расчета модельных сигналов, по которым рассчитывалась матрица Якоби, был использован 2D-фантом *(рис. 9a)*, имитирующий три зуба (Ca, Z = 20), средний зуб с имплантом (Au, Z = 79); синограмма рассчитывалась для 30 кэВ.

Результат реконструкции алгебраическим методом, созданным специально для решения задачи томографии в присутствии сильно поглощающих областей [18] представлен на *рисунке 10a*. Область интересов (region of interest, ROI) приведена на рисунке *10b*.

Сильная неоднородность в распределении Ca внутри зуба с имплантом объясняется наличием сильно поглощающей области импланта внутри зуба. Темная часть на среднем зубе *(рис. 9b)* – результат визуализации матрицы Якоби свидетельствует о том, что уровень доверия результатам реконструкции в этой области низкий.

Заключение

В статье мы попытались проанализировать ошибки, возникающие на этапах формирования томографического изображения. В идеале, если метод компьютерной томографии применяется для метрологических измерений и в руках пользователя есть эталон, о распределении восстанавливаемых характеристик которого информация получена другими методами, то, зная погрешности измерений и сравнивая результаты реконструкции с эталонным распределением характеристик, можно попытаться оценить ошибки, возникшие на этапе реконструкции. Результат будет пригоден только для используемого томографа и для измерений объектов, повторяющих эталон. Это важная задача в области промышленной томографии, цель которой – находить внутренние трещины, характеризовать отклонения размеров скрытых областей и пр. В медицине, в научных исследованиях такой подход не работает по причине отсутствия эталонов. Все, что есть в руках у исследователя – изображения пространства измерений. То есть задача реконструкции – выбрать в пространстве объектов такой объект, для которого отличие результата измерений от результата моделирования, использующего

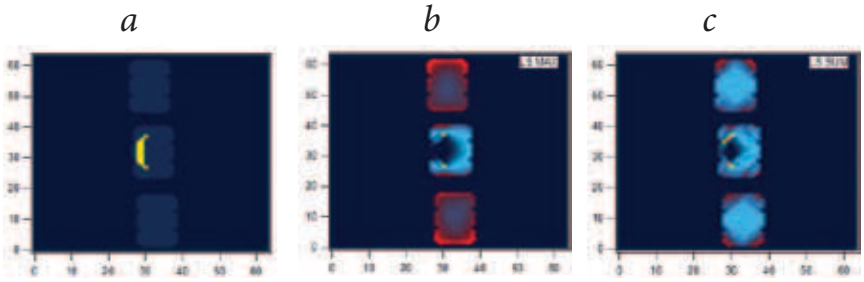


Рис. 9. a – Используемый для моделирования фантом; b, c – два способа представления векторной формы матрицы Якоби.

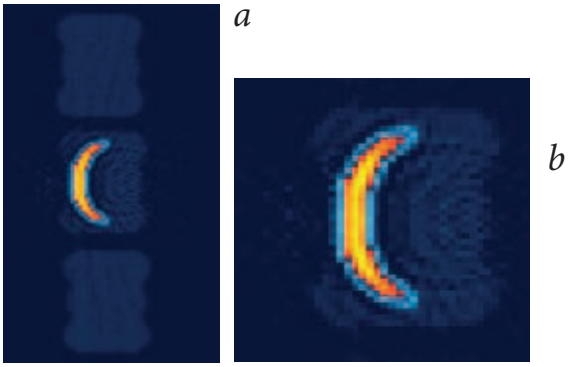


Рис. 10. a – Результат реконструкции; b – увеличенная часть изображения 10a.

полученный результат реконструкции, будет минимальным или максимальной должна быть их схожесть [19]. Финальная ошибка рассчитывается в пространстве измерений следующим образом:

$$E_m = ||H\mu_{opt} - p||, \quad (9)$$

где μ_{opt} – оптимальное изображение из всего пространства изображений, описываемого его свойствами [15]. Ошибка выбора μ_{opt} заложена в процедуре регуляризации на этапе реконструкции.

В дальнейшем мы планируем построить алгоритм автоматического расчета изображения на базе матрицы Якоби и включить матрицу Якоби в детектор схожести изображений в пространстве измерений.

Литература

1. ***В.Е. Асадчиков, В.Г. Бабак, А.В. Бузмаков, Ю.П. Дорохин, И.П. Глаголев, Ю.В. Заневский, В.Н. Зрюев, Ю.С. Кривоносов, В.Ф. Мамич, Л.А. Мосейко, Н.И. Мосейко, Б.В. Мчедлишвили, С.В. Савельев, Р.А. Сенин, Л.П. Смыков, Г.А. Тудоси, В.Д. Фатеев, С.П. Черненко, Г.А. Черемухина, Е.А. Черемухин, А.И. Чуличков, Ю.Н. Шилин, В.А. Шишков***
ПТЭ, 2005, **48**(3), 99.
2. ***И.А. Щелоков, М.В. Чукалина, В.Е. Асадчиков***
Кристаллография, 2015, **60**(5), 673.
DOI: 10.7868/S0023476115050136.
3. ***В.Е. Асадчиков, Р.А. Сенин, А.Е. Благов, А.В. Бузмаков, В.И. Гулимова, Д.А. Золотов, А.С. Орехов, А.С. Осадчая, К.М. Подурец, С.В. Савельев, А.Ю. Серегин, Е.Ю. Терещенко, М.В. Чукалина, М.В. Ковальчук***
Кристаллография, 2012, **57**(5), 782.
4. ***Д.А. Золотов, А.В. Бузмаков, В.Е. Асадчиков, А.Э. Волошин, В.Н. Шкурко, И.С. Смирнов***
Кристаллография, 2011, **56**(3), 426.
5. ***Д.А. Золотов***
Диссерт. канд. физ-мат. наук, Институт кристаллографии им. А.В. Шубникова РАН, Москва, 2011, 132 с.
6. ***C.E. Kahn Jr, J.A. Carrino, M.J. Flynn, D.J. Peck, S.C. Horii***
JASR, 2007, **4**(9), 652. DOI: 10.1016/j.jacr.2007.06.004.
7. *Шкала Хаунсфилда.* (http://doktorland.ru/shkala_haunsfilda.html).
8. ***J. Radon***
Akad. Wiss., 1917, **69**, 262.
9. ***A.C. Kak, M. Slaney***
Principles of Computerized Tomographic Imaging, Ser. Classics in Applied Mathematics, SIAM Publ., 2001, 327 pp.
DOI: 10.1137/1.9780898719277.
10. ***М.В. Чукалина, А.В. Бузмаков, Д.П. Николаев, А.И. Чуличков, М.К. Каримов, Г.А. Расулов, Р.А. Сенин, В.Е. Асадчиков***
Изм. техн., 2008, №2, 19.
11. ***В.Е. Прун, А.В. Бузмаков, Д.П. Николаев, М.В. Чукалина, В.Е. Асадчиков***
Автоматика и телемеханика, 2013, №10, 86.
12. ***D. Nikolaev, S. Karpenko, I. Nikolaev, P. Nikolaev***
В *Proc. ECMS 2008 22nd European Conference on Modelling and Simulation (Cyprus, Nicosia, 3–6 June, 2008)*, ECMS Publ., 2008, pp. 238–243. DOI: 10.7148/2008-0238.
13. ***E. Ershov, A. Terekhin, D. Nikolaev, V. Postnikov, S. Karpenko***
В *Proc. SPIE 9875, Eighth International Conference on Machine Vision*

(ICMV 2015), (Spain, Barcelona, 19–20 November, 2015), SPIE Publ., 2015, 987509. DOI: 10.1117/12.2228852.
14. *Y. Censor, P.P.B. Eggermont, D. Gordon*
Numer. Math., 1983, **41**(1), 83. DOI: 10.1007/BF01396307.
15. *Е.А. Черемухин, А.И. Чуличков*
Ж. вычисл. матем. матем. физ., 2005, **45**(4), 741.
16. *M. Chukalina, D.P. Nikolaev, A. Simionovici*
В *Proc. ECMS 2007 21st European Conference on Modelling and Simulation (Czech Republic, Prague, 4–6 June, 2007),* ECMS Publ., 2008, pp. 309–312. DOI: 10.7148/2007-0309.
17. *J.L. Davidson, C.A. Garcia-Stewart, K.B. Ozanyan, P. Wright, S. Pegrum, H. McCann*
В *Proc. Photon 06 (UK, Manchester, 4–7 September, 2006),* 2006, P2.9, pp. 1–6.
18. *M. Chukalina, D. Nikolaev, V. Sokolov, A. Ingacheva, A. Buzmakov, V. Prun*
В *Proc. SPIE 9875, Eighth International Conference on Machine Vision (ICMV 2015), (Spain, Barcelona, 19 November, 2015),* SPIE Publ., 2015, 98751C. DOI: 10.1117/12.2228810.
19. *D. Tomaževič, B. Likar, F. Pernuš*
Image Anal. Stereol., 2012, **31**(1), 43. DOI: 10.5566/ias.v31.p43-53.

English

The Error Formation in the Computed Tomography: from a Sinogram to the Results Interpretation*

Marina V.Chukalina –
FSRC "Crystallography and Photonics" RAS
59, Leninskiy Ave., Moscow, 119333, Russia
e-mail: chukalinamarina@mail.ru

Dmitriy P. Nikolaev –
A.A. Kharkevich Institute for Information Transmission Problems RAS
19-1, Bolshoy Karetny Per., Moscow, 127051, Russia
e-mail: d.p.nikolaev@gmail.com

Aleksey V. Buzmakov –
FSRC "Crystallography and Photonics" RAS
59, Leninskiy Ave., Moscow, 119333, Russia
e-mail: buzmakov@gmail.com

Anastasiya S. Ingacheva –
FSRC "Crystallography and Photonics" RAS
59, Leninskiy Ave., Moscow, 119333, Russia ,
National Research University "Higher School of Economics"
21/4-1, Staraya Basmannaya Str., Moscow, 105066, Russia
e-mail: ingacheva@gmail.com

Denis A. Zolotov –
FSRC "Crystallography and Photonics" RAS
59, Leninskiy Ave., Moscow, 119333, Russia
e-mail: zolotovden@crys.ras.ru

Andrey P. Gladkov –
A.A. Kharkevich Institute for Information Transmission Problems RAS
19-1, Bolshoy Karetny Per., Moscow, 127051, Russia
e-mail: gladkov.ap@iitp.ru

Victor E. Prun –
Moscow Institute of Physics and Technology (State University)
9, Institutskiy Per., Dolgoprudny, Moscow Region, 141700, Russia
e-mail: vicproon@gmail.com

Boris S. Roshchin –
AFSRC "Crystallography and Photonics" RAS
59, Leninskiy Ave., Moscow, 119333, Russia
e-mail: broshchin@gmail.com

Igor A. Shchelokov –
Institute of Microelectronics Technology and High Purity Materials RAS
6, Academician Ossipyan Str., Chernogolovka, Moscow Region, 142432, Russia
e-mail: igor@iptm.ru

Victoriya I. Gulimova –
Research Institute of Human Morphology
3, Tsyurupy Str., Moscow, 117418, Russia
e-mail: gulimova@yandex.ru

Sergey V. Savelev –
Professor,
Research Institute of Human Morphology
3, Tsyurupy Str., Moscow, 117418, Russia
e-mail: braincase@yandex.ru

Viktor E. Asadchikov –
Professor,
FSRC "Crystallography and Photonics" RAS
59, Leninskiy Ave., Moscow, 119333, Russia
e-mail: asad@ns.crys.ras.ru

Abstract

In this paper for the first time an attempt is made to describe the whole range of error sources that distort the tomographic image during the formation process, to classify the errors sources, and to define qualitatively the connection between causes of the distortions. The use of computed tomography in the areas of medicine and industrial diagnostics makes the cost of a mistake in interpreting its results prohibitively high. Therefore, the final goal of the work is to construct a mathematical equation that relates the magnitude of the final error and parameters of each stage forming the error. All presented in the article tomographic images were obtained by the authors as results of the experiments carried out on a hardware-software complex functioning in the Shubnikov Institute of Crystallography RAS.

Keywords: computed tomography, image formation, sources of the final error formation, errors link analysis.

* *The work was financially supported by RFBR (projects 13-07-00970, 13-07-12179, 13-02-00552 and 16-04-00815).*

Images & Tables

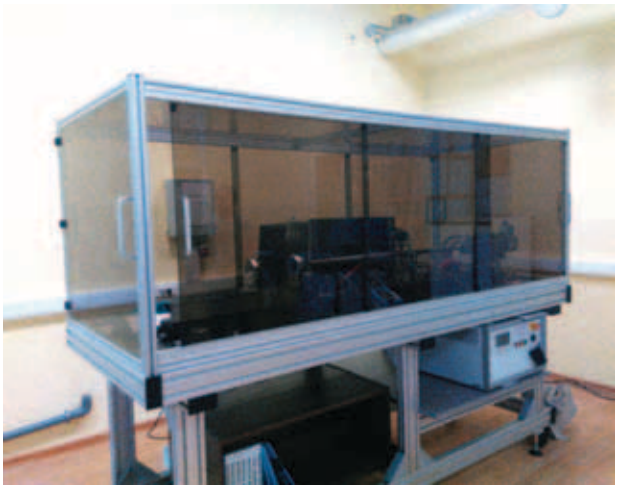


Fig. 1. Hardware-software complex for computed tomography (IC RAS).

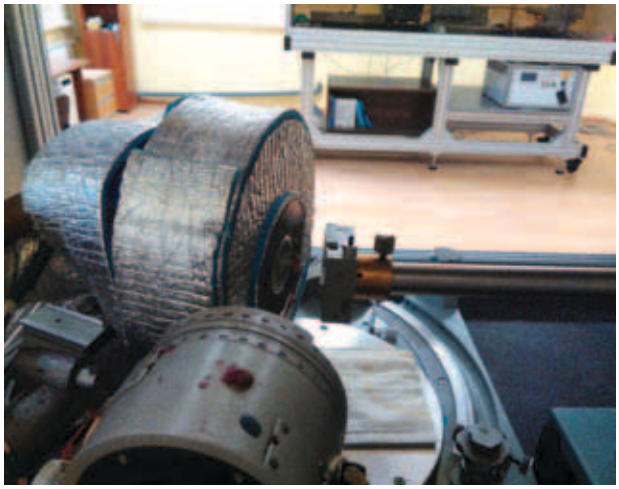


Fig. 4. The unit for topo-tomographic measurements.

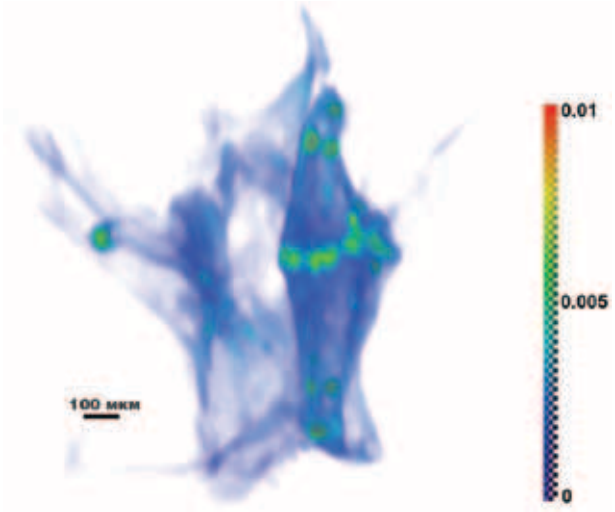


Fig. 2. Result of the tomographic reconstruction of the gecko's vertebra.

Fig. 3. Photo of the natural diamond.

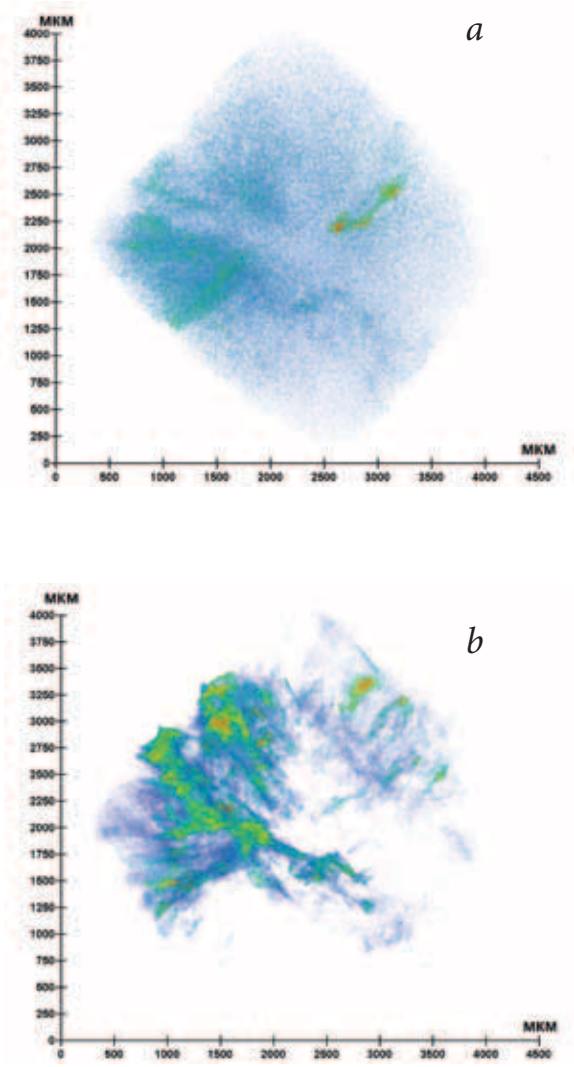


Fig. 5. For the diamond shown in Fig. 3:
a – computed tomography; b – topo-tomography.

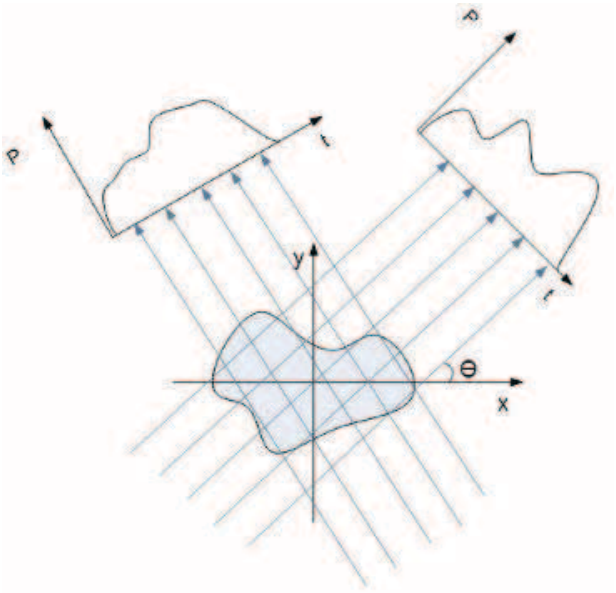


Fig. 6. The parallel measuring scheme.

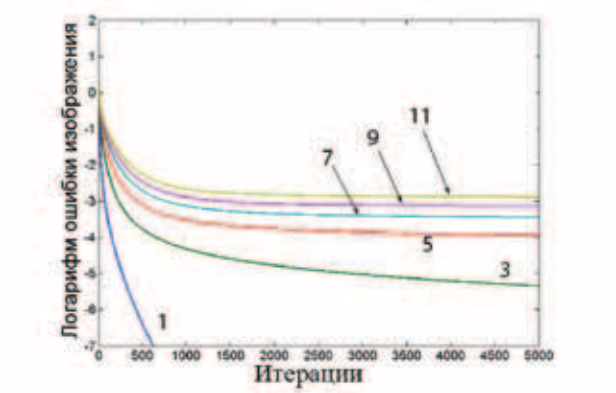


Fig. 7. Dynamics of the root mean square error for algebraic reconstruction technique.

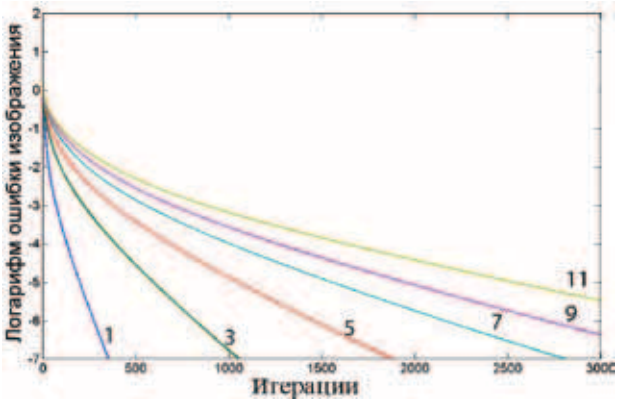


Fig. 8. Dynamics of the root mean square error for algebraic reconstruction technique with subiteration.

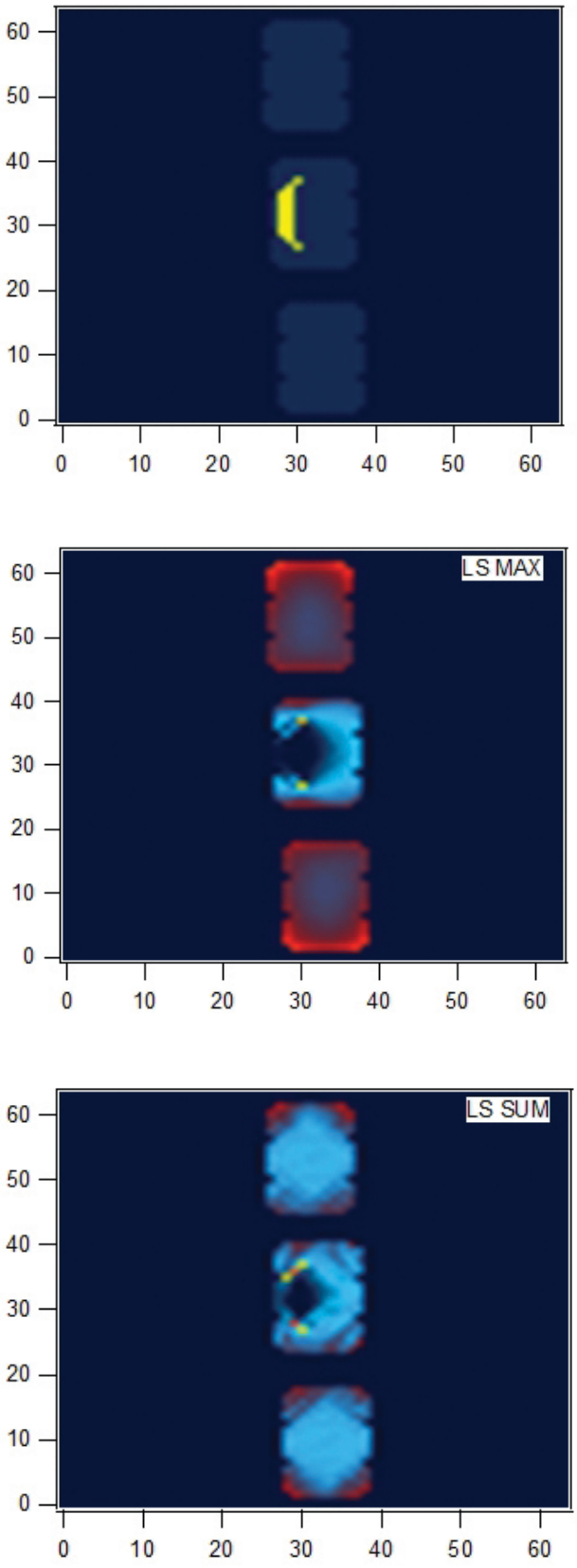


***Fig. 9.** a – Phantom used for simulation; b, c – two ways for visualizing the vector form of the Jacobian matrix.*

DOI: 10.22204/2410-4639-2016-092-04-72-83 № 4 (92) октябрь–декабрь 2016 г.

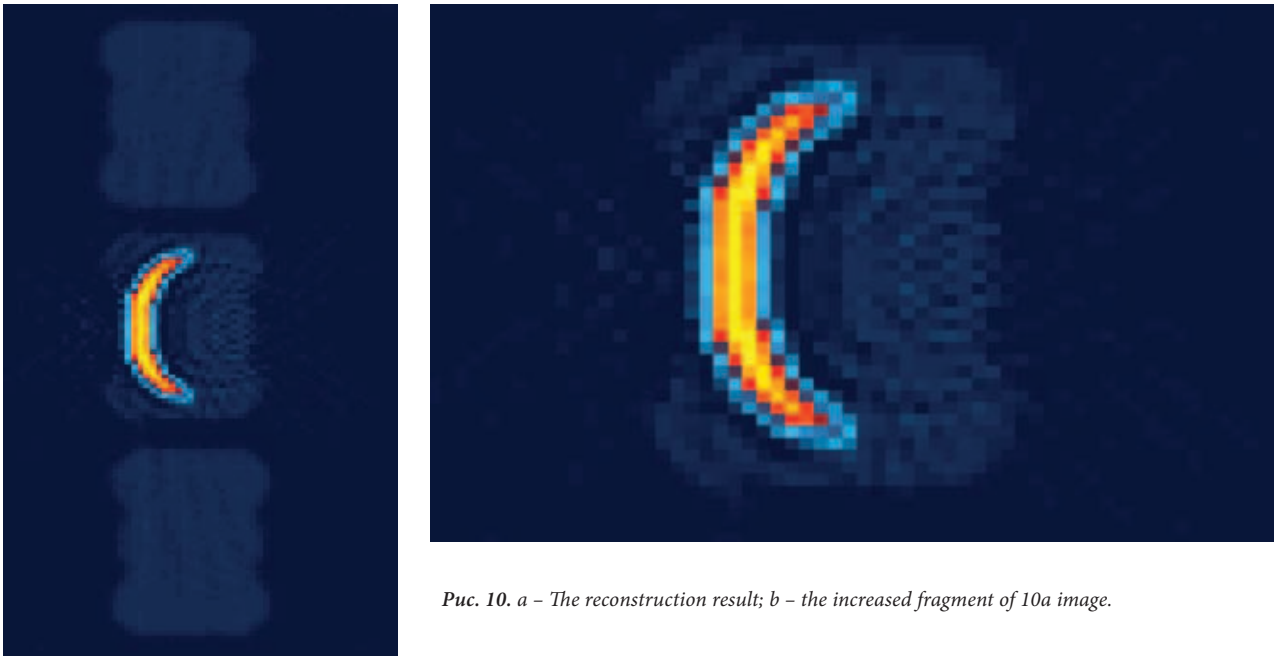


Рис. 10. a – The reconstruction result; b – the increased fragment of 10a image.

References

1. ***V.E. Asadchikov, A.V. Buzmakov, V.N. Zryuev, Yu.S. Krivonsov, B.V. McHedlishvili, R.A. Senin, Yu.N. Shilin, V.A. Shishkov, V.G. Babak, Yu.P. Dorokhin, I.P. Glagolev, V.F. Mamich, V.D. Fateev, Yu.V. Zanevskii, L.P. Smykov, S.P. Chernenko, G.A. Cheremukhina, L.A. Moseiko, N.I. Moseiko, S.V. Savel'ev, R.A. Senin, L.P. Smykov, G.A. Tudosi, V.D. Fateev, S.P. Chernenko, G.A. Cheremukhina, E.A. Cheremukhin, A.I. Chulichkov, Yu.N. Shilin, V.A. Shishkov***
Instrum. Exp. Tech., 2005, **48**(3), 364. DOI: 10.1007/s10786-005-0064-4.
2. ***I.A. Schelokov, M.V. Chukalina, V.E. Asadchikov***
Crystallogr. Rep., 2015, **60**(4), 611. DOI: 10.1134/S1063774515050132.
3. ***V.E. Asadchikov, A.E. Blagov, A.V. Buzmakov, D.A. Zolotov, A.S. Orekhov, A.S. Osadchaya, A.Yu. Seregin, E.Y. Tereshchenko, M.V. Chukalina, M.V. Kovalchuk, V.I. Gulimova, S.V. Savel'ev, R.A. Senin, K.M. Podurets***
Crystallogr. Rep., 2012, **57**(5), 700. DOI: 10.1134/S1063774512050021.
4. ***D.A. Zolotov, A.V. Buzmakov, V.E. Asadchikov, A.E. Voloshin, V.N. Shkurko, I.S. Smirnov***
Crystallogr. Rep., 2011, **56**(3), 393. DOI: 10.1134/S1063774511030345.
5. ***D.A. Zolotov***
PhD Thesis in Phys.-Math. Scien., [Dissertation for the degree of Candidate of Physical-Mathematical Sciences], A.V. Shubnikov Institute of Cristallography RAS, RF, Moscow, 2011, 132 pp. (in Russian).
6. ***C.E. Kahn Jr, J.A. Carrino, M.J. Flynn, D.J. Peck, S.C. Horii***
JASR, 2007, **4**(9), 652. DOI: 10.1016/j.jacr.2007.06.004.
7. *Hounsfield scale [Shkala Khaunsfilda].*
((http://doktorland.ru/shkala_haunsfilda.html) (in Russian).
8. ***J. Radon***
Akad. Wiss., 1917, **69**, 262.
9. ***A.C. Kak, M. Slaney***
Principles of Computerized Tomographic Imaging, Ser. Classics in Applied Mathematics, SIAM Publ., 2001, 327 pp.
DOI: 10.1137/1.9780898719277.
10. ***M.V. Chukalina, A.V. Buzmakov, D.P. Nikolaev, A.I. Chulichkov, M.G. Karimov, G.A. Rasulov, R.A. Senin, V.E. Asadchikov***
Meas. Tech., 2008, **51**(2), 136. DOI: 10.1007/s11018-008-9015-3.
11. ***V.E. Prun, A.V. Buzmakov, D.P. Nikolaev, M.V. Chukalina, V.E. Asadchikov***
Automation and Remote Control, 2013, **74**(10), 1670.
DOI: 10.1134/S000511791310007X.
12. ***D. Nikolaev, S. Karpenko, I. Nikolaev, P. Nikolaev***
In *Proc. ECMS 2008 22nd European Conference on Modelling and Simulation (Cyprus, Nicosia, 3–6 June, 2008)*, ECMS Publ., 2008, pp. 238–243. DOI: 10.7148/2008-0238.
13. ***E. Ershov, A. Terekhin, D. Nikolaev, V. Postnikov, S. Karpenko***
In *Proc. SPIE 9875, Eighth International Conference on Machine Vision (ICMV 2015), (Spain, Barcelona, 19–20 November, 2015)*, SPIE Publ., 2015, 987509. DOI: 10.1117/12.2228852.
14. ***Y. Censor, P.P.B. Eggermont, D. Gordon***
Numer. Math., 1983, **41**(1), 83. DOI: 10.1007/BF01396307.
15. ***E.A. Cheremukhin, A.I. Chulichkov***
Comp. Math. & Math. Phys., 2005, **45**(4), 716.
16. ***M. Chukalina, D.P. Nikolaev, A. Simionovici***
In *Proc. ECMS 2007 21st European Conference on Modelling and Simulation (Czech Republic, Prague, 4–6 June, 2007)*, ECMS Publ., 2008, pp. 309–312. DOI: 10.7148/2007-0309.
17. ***J.L. Davidson, C.A. Garcia-Stewart, K.B. Ozanyan, P. Wright, S. Pegrum, H. McCann***
In *Proc. Photon 06 (UK, Manchester, 4–7 September, 2006)*, 2006, P2.9, pp. 1–6.
18. ***M. Chukalina, D. Nikolaev, V. Sokolov, A. Ingacheva, A. Buzmakov, V. Prun***
In *Proc. SPIE 9875, Eighth International Conference on Machine Vision (ICMV 2015), (Spain, Barcelona, 19 November, 2015)*, SPIE Publ., 2015, 98751C. DOI: 10.1117/12.2228810.
19. ***D. Tomaževič, B. Likar, F. Pernuš***
Image Anal. Stereol., 2012, **31**(1), 43. DOI: 10.5566/ias.v31.p43-53.

Оптимизация быстродействия первых слоев глубоких сверточных нейронных сетей *

Е.Е. Лимонова, А.В. Шешкус, Д.П. Николаев, А.А. Иванова, Д.А. Ильин, В.Л. Арлазаров

В данной работе рассмотрено несколько методов ускорения нейросетевого распознавания образов. В первом методе предложено использование целочисленной арифметики, которая позволила ускорить распознавание на 40% для задач распознавания латинских букв и цифр. Другие два метода модицифируют структуру сверточного слоя нейронной сети с целью уменьшения сложности вычислений. Во втором методе используется представление сверточных фильтров как линейных комбинаций сепарабельных фильтров на этапе создания сети с последующим обучением полученной структуры. Эксперименты показали, что этот метод способен ускорить сверточный слой с 16 фильтрами 11×11 практически в пять раз без потери качества распознавания. Третий метод позволяет сократить количество сверточных фильтров за счет использования перемешивающих сверток, что сохраняет качество распознавания на прежнем уровне и дает ускорение на 10% для сверточного слоя с 16 фильтрами 11×11.

Ключевые слова: сверточные нейронные сети, сложность вычислений, сепарабельные фильтры, перемешивающие свертки.

** Работа выполнена при финансовой поддержке РФФИ (проекты №№ 13-07-12178 и 15-29-06083).*

Введение

Для обработки и анализа данных часто используется такой мощный и современный инструмент, как нейронные сети. Одним из важнейших приложений нейронных сетей является решение задач технического зрения. Например, в рамках технологий технического зрения нейронные сети могут применяться для детектирования, классификации или распознавания объектов.

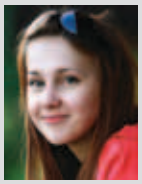
Следует отметить, что все больше задач технического зрения требуют решения в реальном времени, например задачи классификации и детектирования объектов [1–3]. При этом распознавание образов в видеопотоке может требовать достаточно высокой производительности. Кроме того, индустриальные системы часто предъявляют довольно жесткие требования к таким характеристикам используемого оборудования, как производительность, объем памяти и энергопотребление. Все это ведет к необходимости искать способы оптимизации времени распознавания.

Для многих задач распознавания образов используются нейронные сети сверточной архитектуры [4, 5]. Сверточная нейронная сеть состоит из нескольких сверточных слоев с нелинейными функциями активации, за которыми следуют полносвязные слои. На первых сверточных слоях нейронной сети происходит извлечение и обработка первичных признаков, причем именно эти слои часто оказываются крайне трудоемкими вычислительно и их работа занимает значительную часть времени распознавания. При этом методов

ЛИМОНОВА
Елена Евгеньевна
Московский физико-технический институт (государственный университет)

ШЕШКУС
Александр Владимирович
ООО «Смарт Энджинс Сервис»

НИКОЛАЕВ
Дмитрий Петрович
Институт проблем передачи информации им. А.А. Харкевича РАН



ИВАНОВА
Алена Александровна
Институт проблем передачи информации им. А.А. Харкевича РАН

ИЛЬИН
Дмитрий Алексеевич
Институт системного анализа ФИЦ «Информатика и управление» РАН

АРЛАЗАРОВ
Владимир Львович
член-корреспондент РАН, профессор,
Институт системного анализа ФИЦ «Информатика и управление» РАН

ускорения сверточных нейронных сетей не так много, и зачастую не представляется возможным увеличить быстродействие сверточных слоев без изменения характера их работы. Однако такие методы могут вызвать изменение процедуры обработки первичных признаков нейронной сетью, а значит, изменить качество работы распознающей системы.

Относительно хорошо исследованным методом ускорения сверточных нейронных сетей является использование целочисленной арифметики. При таком подходе вычисления в типе данных с плавающей точкой заменяются менее трудоемкими вычислениями в целых числах. Для разных задач распознавания образов можно использовать 8–16-битную арифметику практически без потери точности, как показано в работах [6, 7]. Мы демонстрируем, что в задачах распознавания цифр и латинских букв можно использовать 16-битную целочисленную арифметику при условии того, что применяемые нейронные сети регуляризованы и используют ограниченные функции активации [8].

Кроме того, для оптимизации времени вычислений можно отдельно ускорять самую трудоемкую часть нейронной сети – свертку. Например, вычисление свертки с композицией одномерных фильтров, а не с одним многомерным, имеет меньшую сложность и является более эффективным. Однако чаще всего уже обученные двумерные фильтры являются несепарабельными и их точное представление в виде композиции одномерных фильтров невозможно. Поэтому в ряде существующих методов ускорения свертки используются те или иные способы аппроксимации существующих двумерных фильтров композицией одномерных фильтров [9–11], что час-то приводит к снижению качества распознавания. Другой подход был продемонстрирован в работе [12], где авторы предложили использовать нейронную сеть со специальной структурой свертки, изначально предполагающей сепарабельность сверточных фильтров. На нескольких примерах было показано, что подобная структура нейронной сети может обеспечить качество, сравнимое с качеством исходной сети.

В этой работе мы демонстрируем два метода ускорения нейронной сети путем модификации структуры сверточных слоев на этапе создания сети: использование сепарированной поканальной структуры фильтров и следующих за ними сверток 1×1 и использование сверток 1×1 для увеличения числа выходов свертки параллельно с уменьшением количества сверточных фильтров.

В первом методе ускорения модификация структуры нейронной сети заключается в том, что сверточные фильтры заменяются линейной комбинацией сепарабельных фильтров на этапе создания сети. Для обучения такой нейронной сети можно использовать те же входные данные и алгоритмы обучения, что и для исходной сети. Полученная сеть может давать лучшее качество распознавания, чем исходная с аппроксимацией сверточных фильтров, поскольку они необязательно являются сепарабельными. Нейронную сеть модифицированной структуры можно дополнительно ускорить с помощью, например, целочисленной арифметики.

Второй метод ускорения заключается в том, что мы предлагаем уменьшить количество сверточных фильтров, но сохранить число выходов и объем информации, получаемый в результате свертки, неизменными за счет использования линейных комбинаций выходных значений с помощью сверток 1×1. Обычно свертки 1×1 применяются как средство для снижения числа выходов промежуточных слоев нейронной сети, а также для повышения нелинейности получающейся сети (при использовании вместе с нелинейной функцией активации). Однако наши исследования показали, что вычисление различных линейных комбинаций выходов свертки с последующей нелинейностью повышает качество работы сети, что позволяет уменьшить количество фильтров и повысить производительность.

Архитектура сверточных нейронных сетей

Первая сверточная нейронная сеть была предложена Лекуном [4, 5] и состояла из двух сверточных слоев с субдискретизацией и нескольких полносвязных слоев. Такая структура сети позволила сделать ее реакцию инвариантной к сдвигу, а обработку признаков – одинаковой для разных локальных об-

ластей изображения. На основе этой архитектуры построено множество нейросетевых архитектур, приспособленных для решения конкретных задач, например AlexNet [13], GoogLeNet [14], VGGNet [15].

Рассмотрим сверточный слой нейронной сети. В каждом таком слое выполняется свертка входного изображения с набором сверточных фильтров, добавление смещений и применение нелинейной функции активации. Входное изображение может быть многоканальным, к примеру цветным. Применение одного сверточного фильтра можно описать следующим образом:

$$O(x,y)=\sum_{c}\sum_{\Delta x}\sum_{\Delta y} I(c,x+\Delta x,y+\Delta y)w(c,\Delta x,\Delta y),$$

где (x,y) – точка выходного изображения, O – результат свертки, c – номер канала, I – входное изображение, w – матрица фильтра, а Δx, Δy задают пространственные размеры фильтра *(рис. 1)*. Сам фильтр также можно считать многоканальным, поскольку он содержит разные коэффициенты для разных каналов входного изображения.

Следующим шагом к результату свертки прибавляется смещение и применяется нелинейная функция активации $O'(x,y)=\varphi(O(x,y)+b)$, где O' – выходные значения сверточного слоя для первого фильтра, b – вектор смещения, φ – функция активации. Однако обычно сверточные слои нейронной сети содержат несколько фильтров, поэтому выход сверточного слоя также можно считать многоканальным: по одному каналу от каждого фильтра.

Найдем сложность вычислений в сверточном слое. Пусть $N\times M$ – размер входного изображения, $K\times K$ – размер фильтра, C – число каналов, L – число фильтров. Тогда число умножений, которые составляют основную сложность в слое, будет $O(NMLK^2C)$.

Полносвязные слои нейронной сети можно описать следующим образом: $\vec{y}=\varphi(W\vec{x}+\vec{b})$,

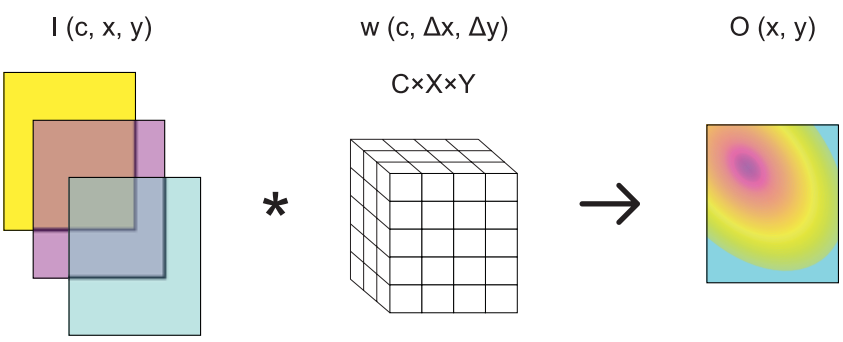


Рис. 1. *Свертка многоканального изображения с фильтром размера C×X×Y.*

где $\vec{x}$ – вход, $\vec{y}$ – выход, W – матрица весов полносвязного слоя, $\vec{b}$ – вектор смещения, φ – нелинейная функция активации.

Использование целочисленной арифметики

Для ускорения вычислений в нейронных сетях можно использовать такие типы данных для весовых коэффициентов, которые обрабатываются процессором быстрее, например целые числа малой разрядности вместо вещественных. Однако при уменьшении разрядности весов (и, как следствие, их точности) изменяются результаты вычислений. Снижение точности вычислений может вести к потере качества распознавания.

Обычно переход от вещественных чисел к целым осуществляется по формуле $x=[f\cdot 2^s]$, где f – вещественное число, x – целое число, являющееся представлением исходного числа f, s – масштабирующий коэффициент. Арифметические операции для целочисленного типа данных можно определить следующим образом:

$$add(x,y)=\min(MaxValue,x+y),$$
$$sub(x,y)=\max(x-y,MinValue),$$
$$div(x,y)=\left[\frac{x}{y}\cdot 2^s\right],$$
$$mul(x,y)=\left[\frac{x\cdot y}{2^s}\right].$$

Источниками неточности вычислений в целых числах являются ошибка при округлении исходных весовых коэффициентов, ошибка округления при выполнении арифметических операций, а также ошибка, возникающая из-за переполнения. Переполнения в нейронной сети можно предотвратить, если использовать ограниченные функции активации. Тогда все про-

DOI: 10.22204/2410-4639-2016-092-04-84-96 № 4 (92) октябрь–декабрь 2016 г.

межуточные результаты остаются ограниченными, поэтому целочисленный тип данных не переполняется и дополнительного снижения точности вычислений не происходит.

Понятно, что весовые коэффициенты и смещения нейронной сети могут оказывать значительное влияние на ошибку вычислений. Их получают в процессе обучения. Основной проблемой при обучении нейронных сетей является переобучение. Переобученная нейронная сеть хорошо работает на примерах из обучающей выборки, однако демонстрирует плохие результаты на примерах, не использовавшихся при обучении. Для предотвращения переобучения применяется регуляризация. Некоторые виды регуляризации, в частности регуляризация по Тихонову, минимизируют абсолютные значения весов, что является очень важным при использовании целочисленной арифметики, поскольку для представления таких весов с достаточно малой потерей точности требуется меньше двоичных разрядов.

Для представления весов нейронной сети мы использовали 16-битный знаковый тип данных с фиксированной точкой: он позволяет ускорить вычисления по сравнению с вещественными 32-битными числами, которые обычно применяются. Например, для процессора Samsung Exynos 5422 с архитектурой ARM использование 16-битных целых чисел позволяет ускорить вычисления в два раза по сравнению с вычислениями в вещественных 32-битных числах. Матричные операции, выполненные с помощью оптимизированной библиотеки линейной алгебры Eigen, также ускорились почти вдвое. Для сверточных нейронных сетей, предназначенных для распознавания латинских букв и цифр и состоящих из пяти слоев (двух сверточных и трех полносвязных), время работы распознавания улучшилось на 40%. Все нейронные сети были регуляризованы для получения максимального качества распознавания.

В *таблице 1* приведены результаты экспериментов. Все нейронные сети были предназначены для распознавания латинских букв. Здесь $\mathrm{ReLU}(x) = \max(0, x)$ – неограниченная функция активации, которая часто используется в сверточных нейронных сетях [13]. Мы заменили ее на ограниченную функцию вида $\mathrm{BReLU}(x) = \min(\max(0, x), a)$, где a – пороговое значение. Fixed – целочисленный тип данных, a float – вещественный 32-битный тип. Полученные результаты показывают, что при отбрасывании дробной части (truncate) качество распознавания уменьшилось, однако при округлении к ближайшему целому (round) и использовании ограниченных функций активации потери качества не произошло.

Таблица 1. *Качество работы нейронных сетей при использовании целочисленной арифметики*

Эксперимент	Параметр				
	Функция активации	Размер тестовой выборки	Качество, %		
			Float	Fixed, truncate	Fixed, round
1	tanh	72965	95.1	69.0	95.3
2	ReLU	45834	96.8	85.2	85.2
3	BReLU	72965	93.3	90.9	93.4

Модификация структуры сверточной нейронной сети

Сепарированная структура сверточного слоя нейронной сети. Будем называть сепарированной структуру сверточного слоя нейронной сети, явно использующую представление сверточных фильтров в виде линейной комбинации сепарабельных фильтров.

В самом простом варианте одномерные фильтры соответствующих размеров просто располагаются последовательно, как показано на *рисунке 2.* Однако этот способ имеет очевидный недостаток: после применения первого фильтра все каналы входного изображения складываются, и большой объем информации теряется. Чтобы преодолеть этот недостаток, мы использовали концепцию разреженной блочной свертки (англ. block sparse convolution), которая представлена cuda-convnet (библиотеке, содержащей инструменты для обучения нейронных сетей с различной структурой [13]) Согласно этой концепции каналы входного изображения делятся на группы, и к каждой группе применяется свой собственный набор фильтров с меньшим числом каналов *(рис. 3).* Таким

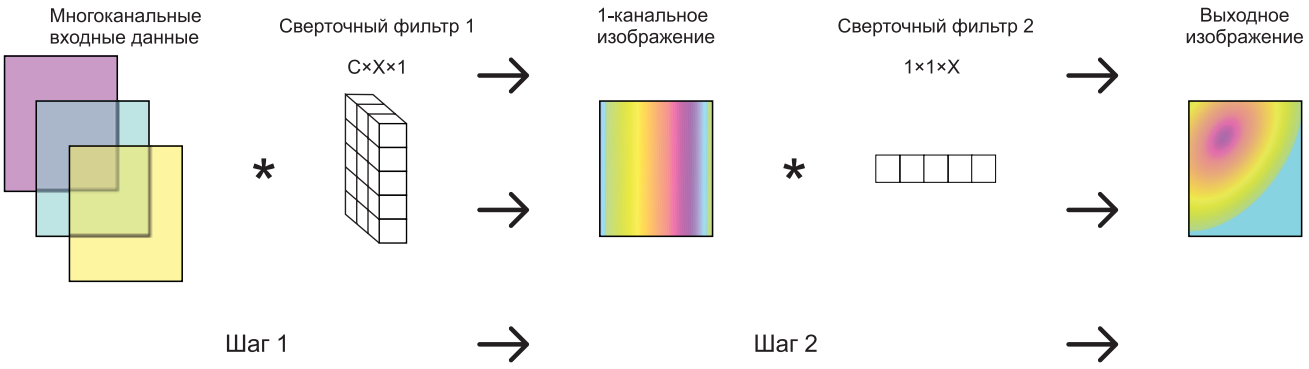


Рис. 2. Простейшая сепарированная структура сверточного слоя нейронной сети.

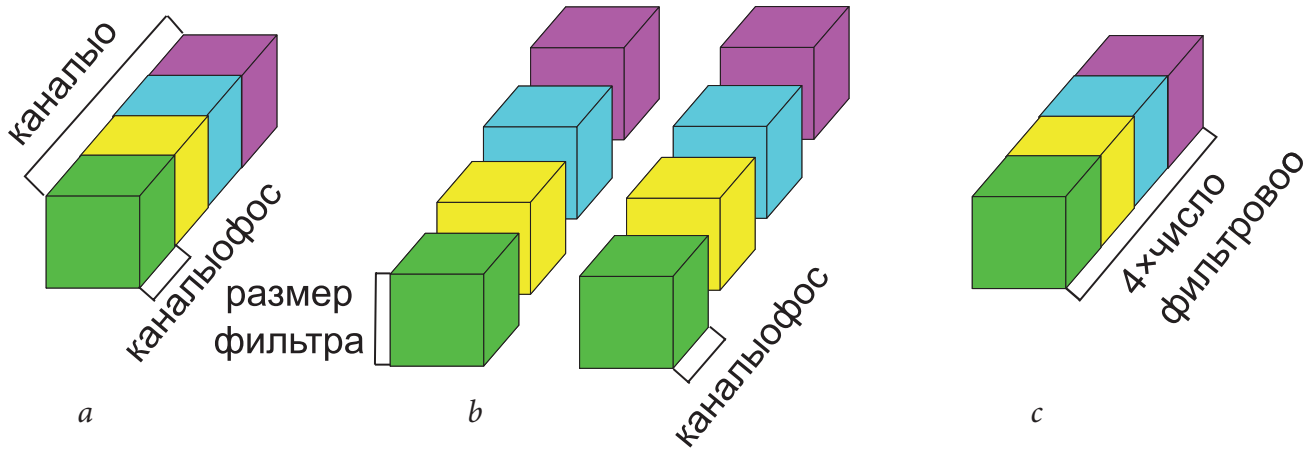


Рис. 3. Слой разреженной блочной свертки с четырьмя группами фильтров по два фильтра в каждой [13]: a – входные каналы разделяются на четыре группы; b – четыре набора фильтров по два фильтра в каждом; c – выход слоя, полученный в результате свертки соответствующих фильтров и входных каналов.

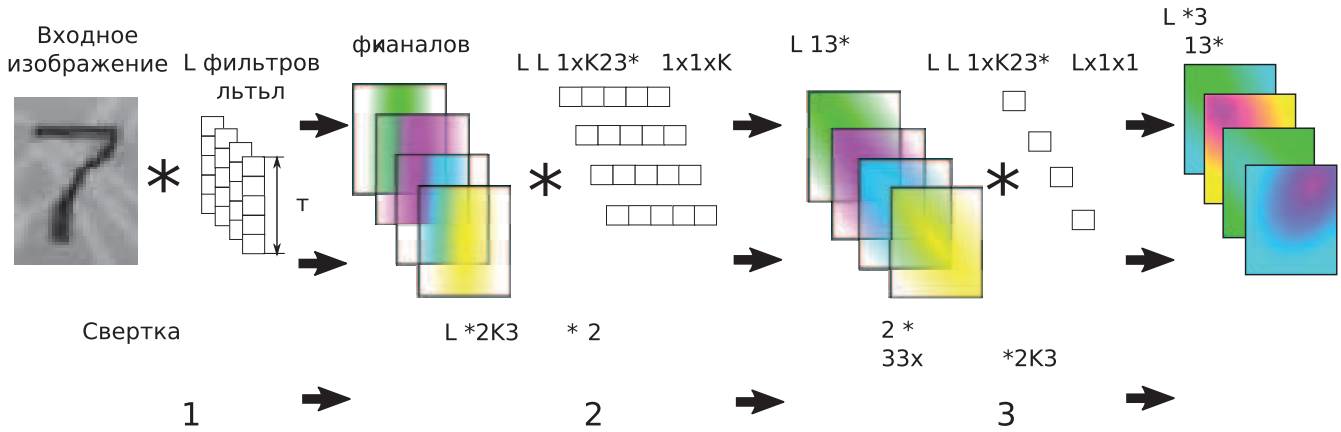


Рис. 4. Предложенная сепарированная структура сверточного слоя с одноканальным входным изображением в качестве примера.

образом, можно организовать различную обработку для разных признаков, а в нашем случае – избежать потери информации внутри измененного сверточного слоя.

Предлагаемая структура свертки показана на *рисунке 4*. Согласно предложенному методу L C-канальных фильтров обычной сверточной сети заменяются структурой из L пар одномерных фильтров и L сверток 1×1, которые нужны для расчета линейных комбинаций, или перемешивания. Вычисления в таком слое будут организованы следующим образом:

1. Применить L C-канальных фильтров размера $K×1$.

2. Разделить каналы на L групп и применить 1-канальный фильтр $1×K$ к каждой группе. Результат вычислений на этом шаге будет содер-

жать L каналов (по одному на каждую группу).

3. Применить L сверток размера 1×1, чтобы получить линейные комбинации выходов шага 2.

4. Применить нелинейную функцию активации.

Сложность вычислений в таком слое будет равна $O(NML(KC+K+L))$, а число весов – $KCL+KL+L^2$. Такая модель содержит меньше весов, чем обычный сверточный слой, а значит, ее избыточность меньше.

Следует отметить, что обучение нейронной сети данной структуры может проводиться стандартными алгоритмами обучения.

Наши эксперименты показывают, что такую структуру сверточного слоя можно успешно использовать для двух задач распознавания образов (*см. далее*). Однако при этом ряд других задач распознавания допускает сепарабельную аппроксимацию фильтров без потери качества, что дает основания полагать, что предложенная структура свертки может быть успешно использована и в них.

Использование перемешивающих сверток 1×1. Выше мы применяли свертки 1×1, чтобы искать линейные комбинации выходов предыдущего шага. Наши эксперименты показали, что в рассмотренных задачах этот шаг необходим для сохранения качества нейронной сети с модифицированной структурой, сравнимого с качеством исходной нейронной сети. При этом мы использовали L сверток 1×1, чтобы не изменять число выходов сверточного слоя и сохранить архитектуру остальной части нейронной сети неизменной.

Рассмотрим следующую структуру сверточного слоя нейронной сети: уменьшим вдвое количество сверточных фильтров и доведем количество выходов слоя до прежнего значения за счет перемешивающих сверток. На примере задач распознавания образов мы продемонстрировали, что использование данной структуры сверточного слоя не вызывает потери качества распознавания. При этом сложность вычисления свертки уменьшается, и система начинает работать более эффективно. Действительно, если мы уменьшим число фильтров до L_1 ($L_1 < L$) и используем L перемешивающих сверток, сложность полученного слоя будет $O(NML_1(K^2C+L))$. При $L_1 < L_1K^2C/(K^2C + L)$ общая сложность вычислений снизится, и мы сможем ускорить работу сети.

Экспериментальные результаты. Мы проводили эксперименты на выборке из русских печатных букв и CIFAR-10. Распознавание выполнялось с помощью сверточных нейронных сетей.

Распознавание русских букв. Обучающая выборка была сформирована из русских букв, вырезанных из фотографий российских паспортов. Она состояла из $9 \cdot 10^5$ изображений размера 20×20 и, помимо букв, содержала точку, тире и пробел *(рис. 5)*. Для валидации использовали по 1024 символа из каждого класса.

Сверточная сеть, на которой проводили эксперименты, состояла из первого сверточного слоя, слоя субдискретизации и нелинейности, второго сверточного слоя и нелинейности, полносвязных слоев и softmax-слоя, который описывается функцией

$$y(j) = \frac{e^{x(j)}}{\sum_{i=1}^{K} e^{x(i)}},$$

где $x(i)$, $i = 1,...,K$ – вход слоя, $y(j)$, $j = 1,...,K$ – выход слоя. В качестве нелинейности использовали ограниченную линейную функцию BReLU. Во время экспериментов изменялся первый сверточный слой, в то время как второй слой оставался неизменным. Он включал в себя 16 фильтров 5×5. Полносвязная часть сети состояла из двух слоев: слоя из 150 нейронов с BReLU и слоя из 36 нейронов.

Рис. 5. *Примеры изображений из обучающей выборки с русскими печатными буквами.*

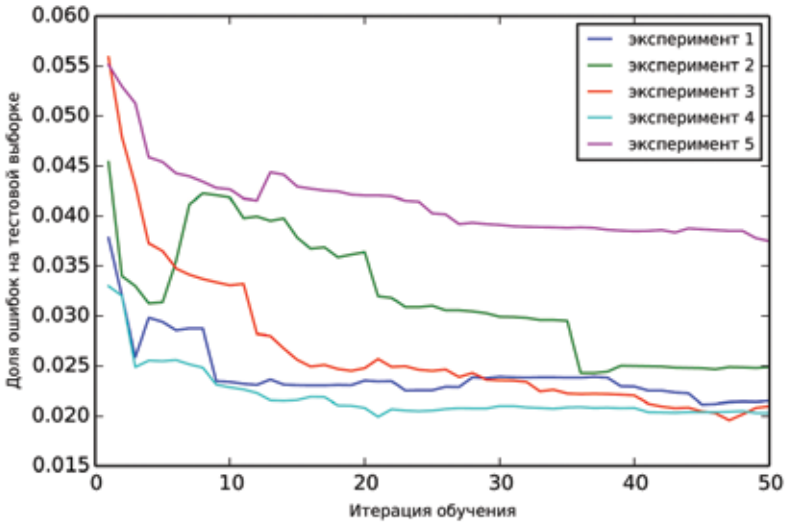


Рис. 6. Ход процесса обучения в экспериментах по распознаванию русских букв.

Таблица 2. *Качество распознавания, обеспечиваемое нейронными сетями с различной структурой свертки при работе с обучающей выборкой из русских букв*

№	Структура свертки	Ошибка , %
1	16×1×5×5	2.1
2	16×1×5×1+16×1×1×5	2.4
3	16×1×5×1+16×1×1×5+16×16×1×1	2.0
4	8×1×5×5+16×8×1×1	2.0
5	8×1×5×5	3.8

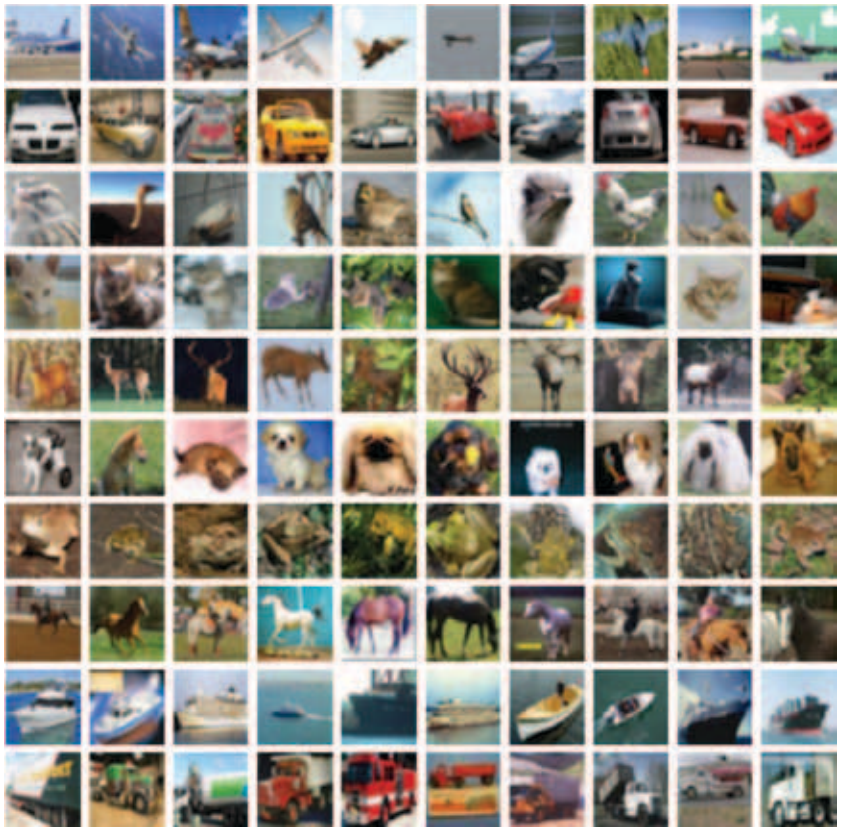


Рис. 7. Примеры изображений из обучающей выборки CIFAR-10.

В общей сложности было проведено пять экспериментов со следующими конфигурациями первого сверточного слоя:

1. 16 одноканальных фильтров 5×5. Этот эксперимент соответствует структуре классической сверточной сети.

2. 16 одноканальных фильтров 5×1, 16 одноканальных фильтров 1×5 с 16 группами. В этом эксперименте мы проверяем необходимость перемешивающих сверток.

3. 16 одноканальных фильтров 5×1, 16 одноканальных фильтров 1×5 с 16 группами, 16 16-канальных фильтров 1×1. Этот эксперимент соответствует предложенной структуре сверточного слоя.

4. Восемь одноканальных фильтров 5×5, 16 16-канальных фильтров 1×1. Здесь мы применяем метод снижения количества фильтров с помощью перемешивающих сверток 1×1.

5. Восемь одноканальных фильтров 5×5. В этом эксперименте продемонстрировано качество работы нейронной сети без перемешивающих сверток.

Процесс обучения показан на *рисунке 6*. Скорости сходимости примерно одинаковы для экспериментов 1, 3 и 4. Это означает, что предложенные методы не вызвали никаких проблем со сходимостью.

Результаты экспериментов представлены в *таблице 2*. Из них следует, что предложенные методы ускорения сверточной нейронной сети не снижают качество работы, что свидетельствует о малой чувствительности нейронных сетей к подобным вариациям первичных признаков в рассмотренной задаче.

CIFAR-10 – открытый набор данных для обучения, состоящий из 60000 цветных изображений 32×32. На этих изображениях были изображены объекты 10 разных классов *(рис. 7)*, по 6000 картинок на класс. Для распознавания этих изображений мы так же обучили сверточную нейронную сеть и провели тот же

набор экспериментов, за тем исключением, что в этой сети мы модифицировали свертку из 64 фильтров 5×5. Процесс обучения показан на *рисунке 8*, а результаты обучения – в *таблице 3*. На основании проведенных экспериментов можно сделать вывод, что и для CIFAR-10 оба предложенных метода не являются причиной снижения качества и возникновения каких-либо проблем при обучении, что также свидетельствует о малой чувствительности нейронных сетей к подобным вариациям первичных признаков и в данной задаче.

Скорость работы системы. Для исследования ускорений, которые можно получить с помощью данных методов, мы измерили время вычислений в сверточном слое. На вычисления в сверточном слое влияет ряд факторов, например размеры входного изображения и размеры фильтров. Это может происходить, поскольку одной из наиболее медленных операций в компьютерных системах является доступ к памяти. Перед любой обработкой данных их необходимо загрузить из памяти. Будем представлять свертку с помощью произведения матриц по методике [16]. Для небольших матриц время инициализации может быть довольно большим по сравнению со временем выполнения умножения. При работе с матрицами большего размера матричное умножение становится вычислительно более трудоемким, и предложенные методы позволяют снизить время работы сети.

В первом эксперименте использовались одно- и трехканальные изображения размера 32×32 и 16 сверточных фильтров переменных размеров. Матричное умножение было реализовано с помощью библиотеки Eigen. Во втором эксперименте мы применяли те же входные изображения и уменьшили число фильтров с 16 до 8. Результаты экспериментов представлены в *таблице 4*.

Таблица 3. *Качество распознавания, обеспечиваемое нейронными сетями с различной структурой свертки при работе с CIFAR-10*

№	Структура свертки	Ошибка , %
1	64×1×5×5	16.5
2	64×1×5×1+64×1×1×5	16.8
3	64×1×5×1+64×1×1×5+64×64×1×1	15.0
4	32×1×5×5+64×32×1×1	14.6
5	32×1×5×5	19.3

Таблица 4. *Время работы сверточных слоев нейронных сетей различной структуры*

Размер фильтров	Время, мкс		
	Классическая структура	Метод 1	Метод 2
Одноканальный вход 32×32			
5×5	65	90	65
7×7	145	100	140
11×11	340	125	320
Трехканальный вход 32×32			
5×5	220	105	205
7×7	425	120	400
11×11	1000	180	895

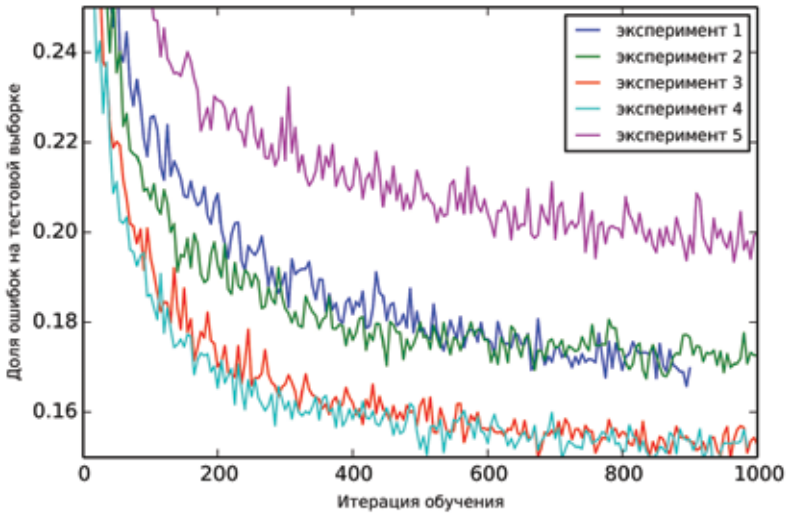


Рис. 8. Ход процесса обучения в экспериментах по распознаванию объектов CIFAR-10.

Очевидно, что для достаточно больших сверточных слоев применение предложенных методов позволяет значительно снизить время работы сверточного слоя нейронной сети.

Заключение

Быстродействие нейронных сетей может иметь принципиальное значение в системах распознавания, поэтому методы их ускорения без потери качества распознавания представляют большой интерес. В работе рассмотрены три метода ускорения нейронных сетей. Первый метод – использование 16-битной целочисленной арифметики – позволил ускорить распознавание на 40% за счет использования более компактного типа данных, который быстрее обрабатывается современными процессорами. Суть двух других методов заключается в снижении сложности вычислений в сверточных нейронных сетях за счет изменения структуры свертки, что делает эти методы подходящими для любых систем распознавания независимо от используемого оборудования. Второй метод заключается в представлении сверточных фильтров в виде линейной комбинации сепарабельных фильтров, а третий – в уменьшении количества сверточных фильтров одновременно с увеличением количества информации, получаемой сверточным слоем, с помощью перемешивающих сверток. Оба этих метода не требуют специального инструментария и модификации процесса обучения. Как показали эксперименты, для сверточных нейронных сетей с относительно большими размерами фильтров или глубоких сверточных сетей эти методы позволяют значительно ускорить работу распознавания без потери качества.

Литература

1. ***E. Kuznetsova, E. Shvets, D. Nikolaev***
   В *Proc. SPIE 9875, Eighth International Conference on Machine Vision (ICMV 2015), (Spain, Barselona, 19–20 November, 2015),* US, Bellingham, SPIE Publ., 2015, 98750N. DOI: 10.1117/12.2228707.
2. ***A. Mastov, I. Konovalenko, A. Grigoryev***
   В *Proc. SPIE 9875, Eighth International Conference on Machine Vision (ICMV 2015), (Spain, Barselona, 19–20 November, 2015),* US, Bellingham, SPIE Publ., 2015, 98750M. DOI: 10.1117/12.2228623.
3. ***V.N. Kopenkov, V.V. Myasnikov***
   В *Posters Proc. 23rd Int. Conf. Computer Graphics, Visualization and Computer Vision (WSCG 2015), (Czech Republic, Plzen, 8–12 June, 2015),* Czech Republic, Plzen, Publ. Vaclav Skala – UNION Agency, 2015, pp. 65–68.
4. ***Y. Lecun, L. Bottou, Y. Bengio, P. Haffner***
   Proc. IEEE, 1998, **86**(11), 2278. DOI: 10.1109/5.726791.
5. ***Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel***
   В *Advances in Neural Information Processing Systems 2 (Proc. NIPS 1989), (USA, CO, Denver, 27–30 November, 1989),* USA, MA, Cambridge, The MIT Press, 1990, pp. 396–404.
6. ***V. Vanhoucke, A. Senior, M.Z. Mao***
   В *Proc. NIPS 2011 Workshop on Deep Learning and Unsupervised Feature Learning, (Spain, Granada, 16 December 2011),* pp. 1–8. (https://deeplearningworkshopnips2011.files.wordpress.com/2011/12/8.pdf).
7. ***S. Gupta, A. Agrawal, K. Gopalakrishnan, P. Narayanan***
   В *JMLR W&CP, Vol. 37: Proceedings of The 32nd International Conference on Machine Learning (France, Lille, 6–11 July 2015),* France, Lille, Microtome Publ., 2015, pp. 1737–1746.
8. ***E. Limonova, D. Ilin, D. Nikolaev***
   В *Proc. SPIE 9875, Eighth International Conference on Machine Vision (ICMV 2015),* (*Spain, Barselona, 19–20 November, 2015*), US, Bellingham, SPIE Publ., 2015, 98750L. DOI: 10.1117/12.2228594.
9. ***R. Rigamonti, A. Sironi, V. Lepetit, P. Fua***
   В *Proc. 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (USA, OR, Portland, 23–28 June 2013),* USA, Md., Silver Spring, IEEE Computer Society Press, 2013, pp. 2754–2761. DOI: 10.1109/CVPR.2013.355.
10. ***E.L. Denton, W. Zaremba, J. Bruna, Y. LeCun, R. Fergus***
    В *Advances in Neural Information Processing Systems 27 (Proc. NIPS 2014), (Canada, Montreal, 8–13 December, 2014),* USA, NY, Red Hook, Publ. Curran Associates, Inc., 2014, pp. 1269–1277.
11. ***M. Jaderberg, A. Vedaldi, A. Zisserman***
    В *Proc. British Machine Vision Conference 2014, (UK, Nottingham, 1–5 September, 2014),* UK, Nottingham, BMVA Press, 2014, p. 55. DOI: 10.5244/C.28.88.
12. ***J. Jin, A. Dundar, E. Culurciello***
    CoRR, 2014, abs/1412.5474. (http://arxiv.org/pdf/1412.5474v4.pdf).
13. ***A. Krizhevsky, I. Sutskever, G.E. Hinton***
    В *Advances in Neural Information Processing Systems 25 (Proc. NIPS 2012), (USA, NV, Stateline, 3–8 December, 2012),* USA, NY, Red Hook, Publ. Curran Associates, Inc., 2012, pp. 1097–1105.
14. ***C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich***
    В *Proc. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (USA, MA, Boston, 7–12 June, 2015),* USA, MD, Silver Spring, IEEE Computer Society Press, 2015, pp. 1–9. DOI: 10.1109/CVPR.2015.7298594.
15. ***K. Simonyan, A. Zisserman***
    CoRR, 2015, abs/1409.1556. (https://arxiv.org/pdf/1409.1556.pdf).
16. ***K. Chellapilla, S. Puri, P. Simard***
    В *Proc. Tenth International Workshop on Frontiers in Handwriting Recognition, (France, La Baule, 23–26 October, 2006),* France, La Baule, Suvisoft, 2006, p. 1038.

English

Performance Optimization of the Initial Layers of Deep Convolutional Neural Networks*

Elena E. Limonova –
Moscow Institute of Physics and Technology (State University)
9, Institutskiy Per., Dolgoprudny, Moscow Region, 141700, Russia
e-mail: elena.e.limonova@gmail.com

Aleksandr V. Sheshkus –
Smart Engines Service Ltd.
9, 60-letiya Octyabrya Ave., Moscow, 117312, Russia
e-mail: astdcall@gmail.com

Dmitriy P. Nikolaev –
A.A. Kharkevich Institute for Information Transmission Problems RAS
19-1, Bolshoy Karetny Per., Moscow, 127051, Russia
e-mail: d.p.nikolaev@gmail.com

Alena A. Ivanova –
A.A. Kharkevich Institute for Information Transmission Problems RAS
19-1, Bolshoy Karetny Per., Moscow, 127051, Russia
e-mail: ivanova@iitp.ru

Dmitriy A. Ilin –
Institute for System Analysis FRC "Computer Science and Control" RAS
9, 60-letiya Octyabrya Ave., Moscow, 117312, Russia
e-mail: dmitry.ilin@phystech.edu

Vladimir L. Arlazarov –
RAS Corresponding Member, Professor,
Institute for System Analysis FRC "Computer Science and Control" RAS
9, 60-letiya Octyabrya Ave., Moscow, 117312, Russia
e-mail: vladimir.arlazarov@gmail.com

Abstract

In this investigation the authors considered several methods of acceleration of images recognition by neural networks. The first method proposes to use the fixed-point arithmetic that allows accelerating the Latin letters and numerals recognition up to 40%. The other two methods involve modification of the structure of a neural network convolutional layer in order to reduce the computational complexity. The second method uses the representation of the convolutional filters as a linear combination of separable filters at the stage of the neural network formation, and, thereafter, a training process of the resultant structure occurs. The experiments showed a fivefold acceleration of the convolutional layer, consisting of 16 filters with 11x11 dimensions, without loss of recognition quality. Finally, the third method reduces the number of convolutional filters and increases the number of convolutional outputs by means of fusing convolutions. Fusing convolutions allow us to maintain the recognition accuracy level, while the processing time of convolutional layer, consisting of 16 filters with 11x11 dimensions, is decreased by 10%.

Keywords: convolutional neural networks, computational complexity, separable filters, fusing convolutions.

Images & Tables

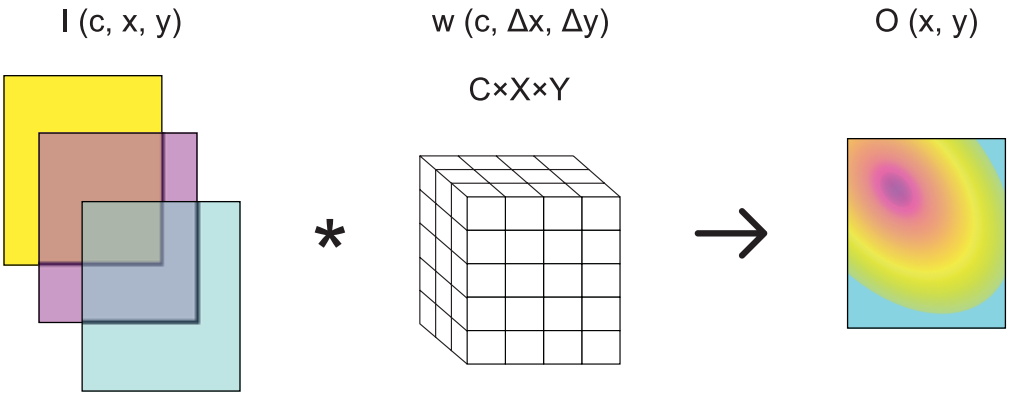


Fig. 1. Convolution of multichannel image with one C×X×Y filter.

* *The work was financially supported by RFBR (projects 13-07-12178 and 15-29-06083).*

Table 1. *Accuracy of different neural networks using fixed-point arithmetic*

Experiment	Parameter				
	Activation function	Test dataset size	Accuracy, %		
			Float	Fixed, truncate	Fixed, round
1	tanh	72965	95.1	69.0	95.3
2	ReLU	45834	96.8	85.2	85.2
3	BReLU	72965	93.3	90.9	93.4

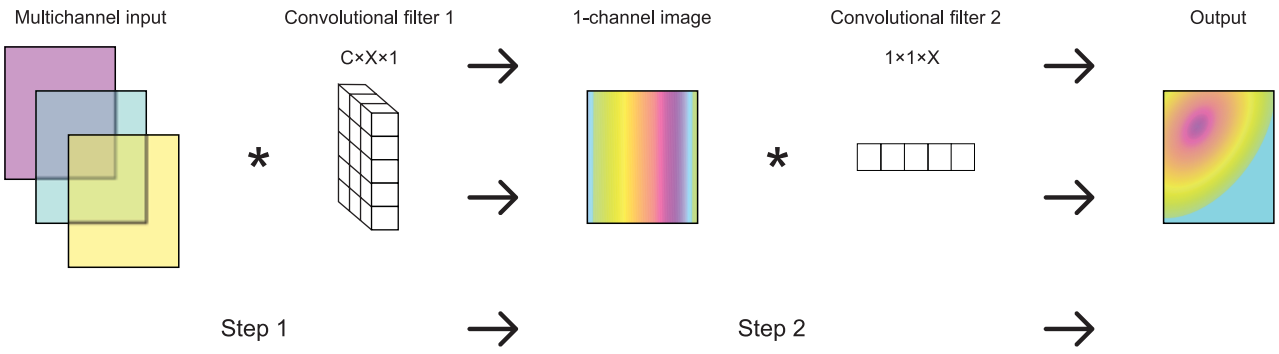


Fig. 2. Simple separated structure of convolutional layer of neural network.

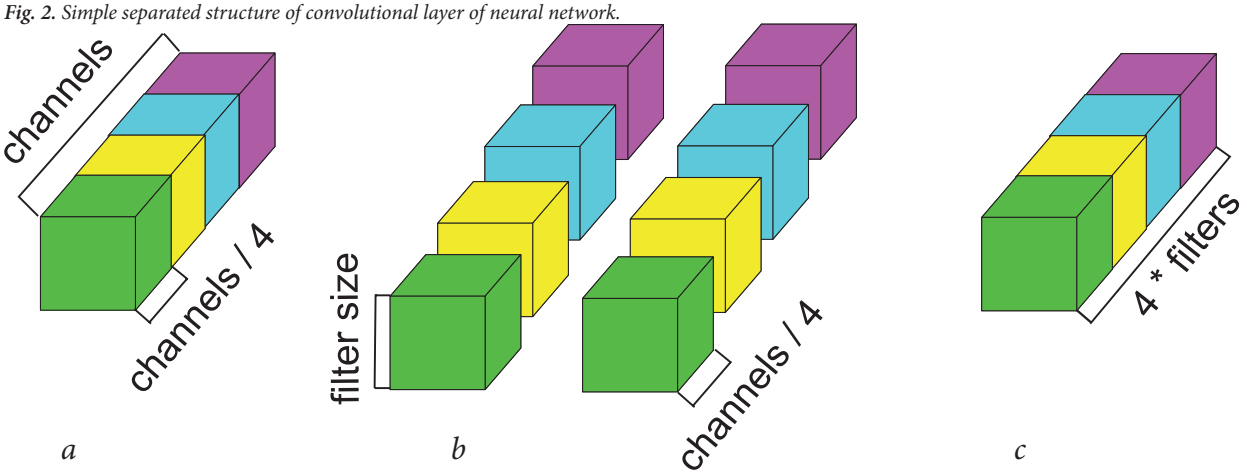


Fig. 3. *Block sparse convolutional layer with 4 groups and 2 filters in each group [13]: a – input channels divided into 4 groups, b – 4 sets of filters with 2 filters in each one, c – output produced by convolution of corresponding filters and input channels.*

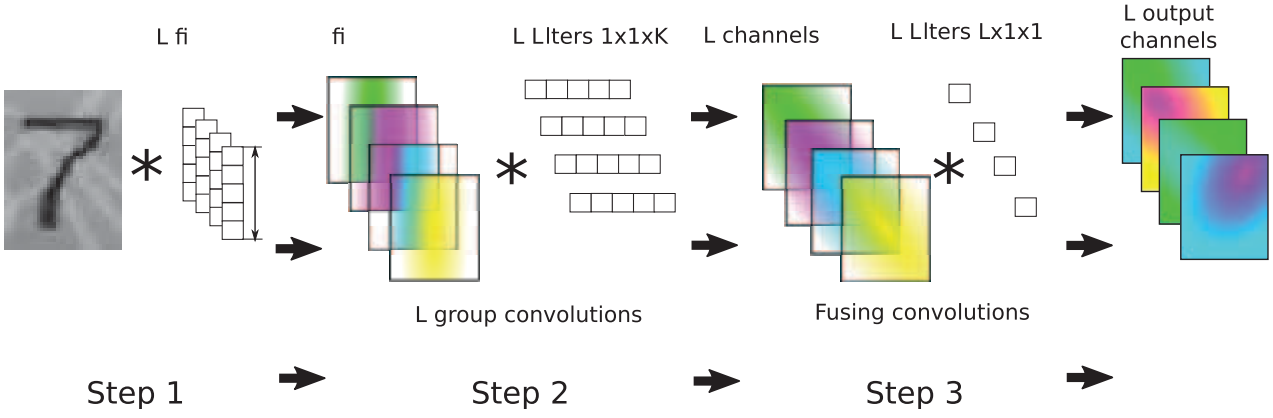


Fig. 4. Proposed separated structure of convolutional layer in the case of 1-channel input.

Fig. 5. Samples of images from Russian printed letters dataset.

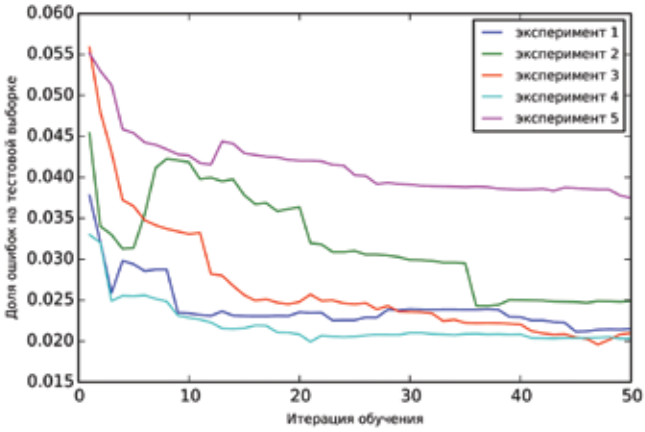


Fig. 6. Training process in experiments on Russian letters dataset.

Table 2. *Recognition accuracy provided by neural networks with different convolution structure on Russian letters training dataset*

№	Convolution structure	Error rate, %
1	16×1×5×5	2.1
2	16×1×5×1+16×1×1×5	2.4
3	16×1×5×1+16×1×1×5+16×16×1×1	2.0
4	8×1×5×5+16×8×1×1	2.0
5	8×1×5×5	3.8

Table 3. *Recognition accuracy provided by neural networks with different convolution structure with CIFAR-10 training dataset*

№	Convolution structure	Error rate, %
1	64×1×5×5	16.5
2	64×1×5×1+64×1×1×5	16.8
3	64×1×5×1+64×1×1×5+64×64×1×1	15.0
4	32×1×5×5+64×32×1×1	14.6
5	32×1×5×5	19.3

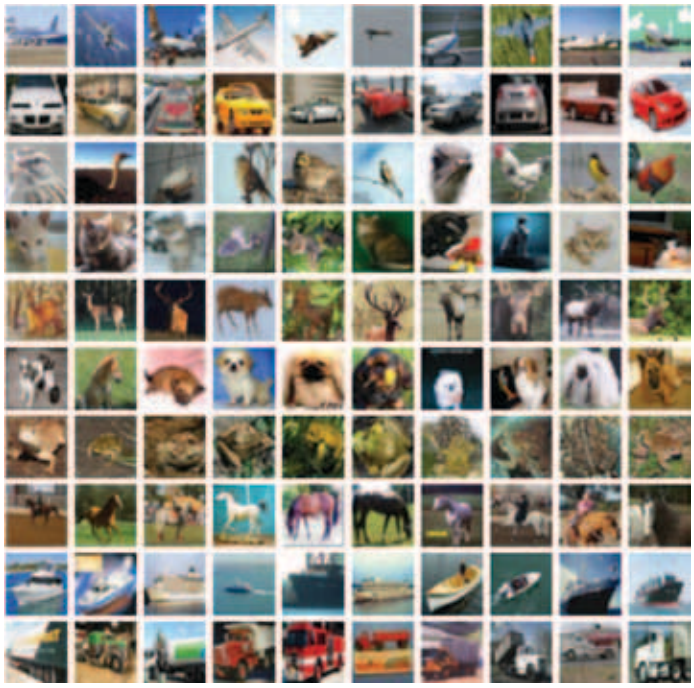


Fig. 7. Samples of images from CIFAR-10 dataset.

Table 4. *Time measurements of the operation of convolutional layers of different neural networks*

Filter size	Time, μs		
	Classical structure	Method 1	Method 2
1-Channel input 32×32			
5×5	65	90	65
7×7	145	100	140
11×11	340	125	320
3-Channels input 32×32			
5×5	220	105	205
7×7	425	120	400
11×11	1000	180	895

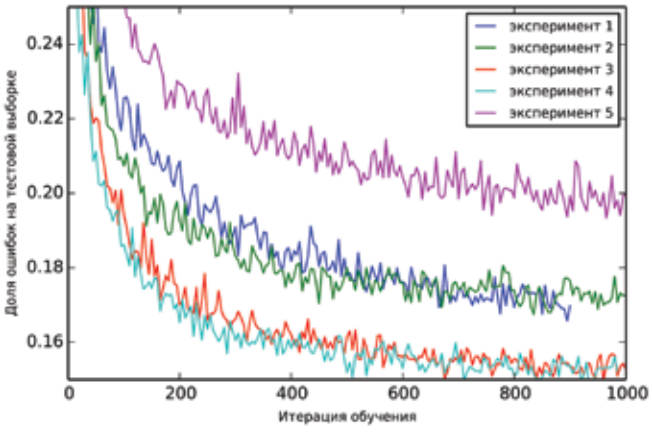


Fig. 8. Training process in experiments on CIFAR-10 dataset.

References

1. ***E. Kuznetsova, E. Shvets, D. Nikolaev***
In *Proc. SPIE 9875, Eighth International Conference on Machine Vision (ICMV 2015), (Spain, Barselona, 19–20 November, 2015),* US, Bellingham, SPIE Publ., 2015, 98750N. DOI: 10.1117/12.2228707.
2. ***A. Mastov, I. Konovalenko, A. Grigoryev***
In *Proc. SPIE 9875, Eighth International Conference on Machine Vision (ICMV 2015), (Spain, Barselona, 19–20 November, 2015),* US, Bellingham, SPIE Publ., 2015, 98750M. DOI: 10.1117/12.2228623.
3. ***V.N. Kopenkov, V.V. Myasnikov***
In *Posters Proc. 23rd Int. Conf. Computer Graphics, Visualization and Computer Vision (WSCG 2015), (Czech Republic, Plzen, 8–12 June, 2015),* Czech Republic, Plzen, Publ. Vaclav Skala – UNION Agency, 2015, pp. 65–68.
4. ***Y. Lecun, L. Bottou, Y. Bengio, P. Haffner***
Proc. IEEE, 1998, **86**(11), 2278. DOI: 10.1109/5.726791.
5. ***Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel***
In *Advances in Neural Information Processing Systems 2 (Proc. NIPS 1989), (USA, CO, Denver, 27–30 November, 1989),* USA, MA, Cambridge, The MIT Press, 1990, pp. 396–404.
6. ***V. Vanhoucke, A. Senior, M.Z. Mao***
In *Proc. NIPS 2011 Workshop on Deep Learning and Unsupervised Feature Learning, (Spain, Granada, 16 December 2011),* pp. 1–8. (https://deeplearningworkshopnips2011.files.wordpress.com/2011/12/8.pdf).
7. ***S. Gupta, A. Agrawal, K. Gopalakrishnan, P. Narayanan***
In *JMLR W&CP, Vol. 37: Proceedings of The 32nd International Conference on Machine Learning (France, Lille, 6–11 July 2015)*, France, Lille, Microtome Publ., 2015, pp. 1737–1746.
8. ***E. Limonova, D. Ilin, D. Nikolaev***
In *Proc. SPIE 9875, Eighth International Conference on Machine Vision (ICMV 2015), (Spain, Barselona, 19–20 November, 2015)*, US, Bellingham, SPIE Publ., 2015, 98750L. DOI: 10.1117/12.2228594.
9. ***R. Rigamonti, A. Sironi, V. Lepetit, P. Fua***
In *Proc. 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (USA, OR, Portland, 23–28 June 2013),* USA, Md., Silver Spring, IEEE Computer Society Press, 2013, pp. 2754–2761. DOI: 10.1109/CVPR.2013.355.
10. ***E.L. Denton, W. Zaremba, J. Bruna, Y. LeCun, R. Fergus***
In *Advances in Neural Information Processing Systems 27 (Proc. NIPS 2014), (Canada, Montreal, 8–13 December, 2014),* USA, NY, Red Hook, Publ. Curran Associates, Inc., 2014, pp. 1269–1277.
11. ***M. Jaderberg, A. Vedaldi, A. Zisserman***
In *Proc. British Machine Vision Conference 2014, (UK, Nottingham, 1–5 September, 2014),* UK, Nottingham, BMVA Press, 2014, p. 55. DOI: 10.5244/C.28.88.
12. ***J. Jin, A. Dundar, E. Culurciello***
CoRR, 2014, abs/1412.5474. (http://arxiv.org/pdf/1412.5474v4.pdf).
13. ***A. Krizhevsky, I. Sutskever, G.E. Hinton***
In *Advances in Neural Information Processing Systems 25 (Proc. NIPS 2012), (USA, NV, Stateline, 3–8 December, 2012),* USA, NY, Red Hook, Publ. Curran Associates, Inc., 2012, pp. 1097–1105.
14. ***C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich***
In *Proc. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (USA, MA, Boston, 7–12 June, 2015),* USA, MD, Silver Spring, IEEE Computer Society Press, 2015, pp. 1–9. DOI: 10.1109/CVPR.2015.7298594.
15. ***K. Simonyan, A. Zisserman***
CoRR, 2015, abs/1409.1556. (https://arxiv.org/pdf/1409.1556.pdf).
16. ***K. Chellapilla, S. Puri, P. Simard***
In *Proc. Tenth International Workshop on Frontiers in Handwriting Recognition, (France, La Baule, 23–26 October, 2006),* France, La Baule, Suvisoft, 2006, p. 1038.

Ключевые аспекты распознавания документов с использованием малоразмерных цифровых камер*

Д.В. Полевой, К.Б. Булатов, Н.С. Скорюкина, Т.С. Чернов, В.В. Арлазаров, А.В. Шешкус

В статье обобщен цикл исследований, посвященных решению задачи распознавания документов, удостоверяющих личность, с помощью малоразмерных цифровых камер. Приведены основные отличительные особенности документов, процесса съемки и их влияние на распознавание. Представлен новый подход к построению системы распознавания как системы с обратной связью.

Ключевые слова: распознавание документов, оптическое распознавание текста на устройстве, межкадровая интеграция.

** Работа выполнена при финансовой поддержке РФФИ (проекты №№ 13-07-12171, 13-07-12172 и 14-07-00730).*

Введение

За последние несколько десятилетий классическое полнотекстовое распознавание превратилось из научной задачи в стандартную функцию информационных систем, которые работают с текстовыми документами. При этом классический подход предполагает использование сканера, планшетного или протяжного, в качестве источника единичного изображения. В то же время в современном мире наблюдается серьезная тенденция к решению различных прикладных задач с использованием стационарных и мобильных малоразмерных цифровых видеокамер, к которым можно в первую очередь отнести web-камеры и камеры мобильных устройств. Серьезный прогресс как в области роста вычислительных ресурсов, так и в области повышения качества и разрешения фото- и видеосъемки дает возможность рассматривать такие устройства в качестве платформы для решения различных задач распознавания, в том числе и распознавания структурированных документов, важным случаем которых является документ, удостоверяющий личность. В рамках работ по проектам РФФИ №№ 13-07-12171-офи_м, 13-07-12172-офи_м и 14-07-00730 нами были исследованы различные аспекты распознавания документов, удостоверяющих личность, в видеопотоке, полученном с мобильных устройств. Наиболее интересные результаты представлены далее.

Разнообразие документов

Наиболее стандартизованной частью современных документов можно считать машиночитаемые зоны, содержание и визуальное оформление которых формализовано и регулируется международными и локальными нормативными документами [1]. С точки зрения оптического распознавания требования к характеристикам машиночитаемых

ПОЛЕВОЙ Дмитрий Валерьевич
Национальный исследовательский технологический университет МИСиС

БУЛАТОВ Константин Булатович
Национальный исследовательский технологический университет МИСиС

СКОРЮКИНА Наталья Сергеевна
Национальный исследовательский технологический университет МИСиС

ЧЕРНОВ Тимофей Сергеевич
Национальный исследовательский технологический университет МИСиС

АРЛАЗАРОВ Владимир Викторович
Московский физико-технический институт (государственный университет)

ШЕШКУС Александр Владимирович
ООО «Смарт Энджинс Сервис»

Рис. 1. Фрагменты изображений документов с различными типами фона и заполнения реквизитов.

зон оптимизированы для получения результатов с использованием специализированного сканирующего оборудования, работающего в инфракрасном диапазоне. На практике для печати фонового заполнения машиночитаемой зоны могут использоваться краски, которые «прозрачны» в инфракрасном диапазоне и в то же время довольно «плотны» в оптическом *(рис. 1)*.

Для одного типа документов может существовать много различных вариантов бланка; шрифтовое оформление бывает весьма разнообразным, причем часть документов может содержать поля с рукописным заполнением. Поэтому система распознавания должна предусматривать возможность отказа как в случае рукописного заполнения документа, так и в случае неприемлемых с точки зрения качества условий съемки.

Получение изображений

Основной особенностью получения изображений документов при помощи малоформатных цифровых камер [2] является невозможность контролировать условия съемки: изображение документа проективно искажается, может происходить

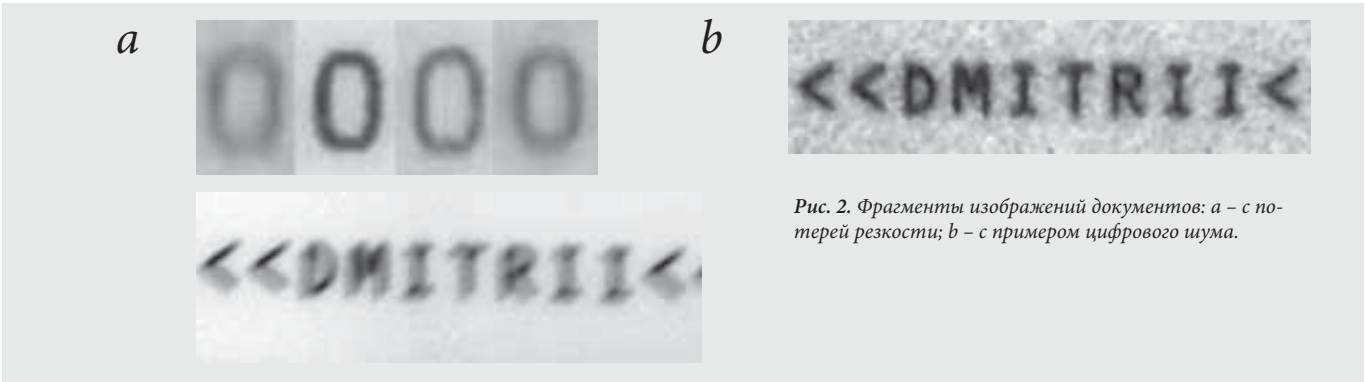


Рис. 2. Фрагменты изображений документов: a – с потерей резкости; b – с примером цифрового шума.

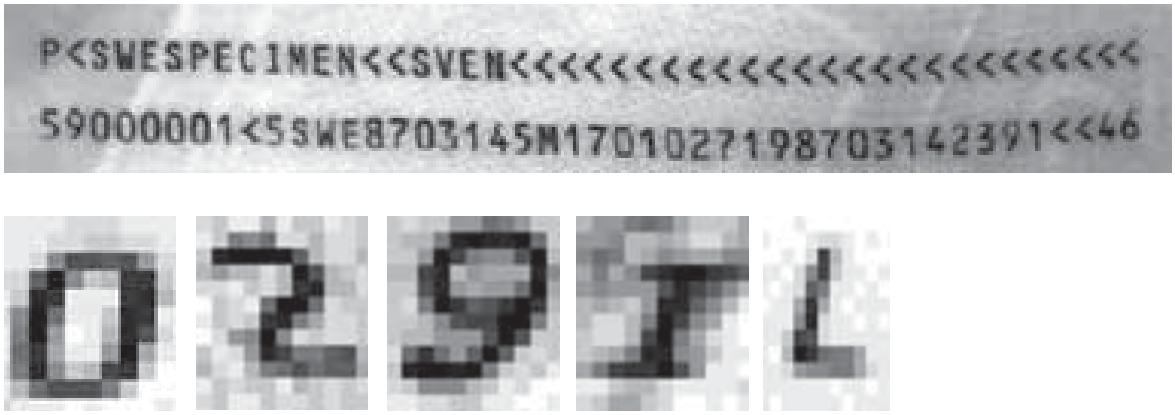


Рис. 3. Фрагменты изображений документов с нелинейными искажениями строк и символов.

 DOI: 10.22204/2410-4639-2016-092-04-97-108 № 4 (92) октябрь–декабрь 2016 г.

Рис. 4. Фрагменты изображений защищенных документов.

«размытие» и «смазывание» *(рис. 2a)*, возможно появление заметных проявлений «цифрового шума» *(рис. 2b)* и артефактов сжатия.

При съемке камерой обеспечить отсутствие механических деформаций документа сложно или просто невозможно, что в комбинации с проективным искажением приводит к существенному искажению изображений строк и символов даже после проективной нормализации *(рис. 3)*.

Используемые при изготовлении документов технологии защиты от подделки в процессе съемки могут вносить дополнительные искажения. Например, пластиковое покрытие или ламинирование бланка документа часто приводят к появлению бликов и отражений *(рис. 4)*, а ламинирующая пленка может иметь голографические изображения с эффектом движения (кинеграммы).

Таким образом, современные малоформатные цифровые камеры позволяют получать изображения документов достаточного для распознавания качества, но в неконтролируемых условиях съемки всегда присутствует искажения, которые необходимо учитывать.

Архитектура обработки видеопотока

Полученное с помощью малоразмерной цифровой камеры изображение документа (обычно одно) может распознаваться на удаленном сервере с использованием клиент-серверной или облачной архитектуры системы. При этом возникают как чисто технические проблемы, так и проблемы соблюдения правил обработки персональной информации. К техническим проблемам необходимо отнести вопросы достаточной скорости, надежности и безопасности передачи данных видеопотока. Для фотографий главной проблемой является то, что результат не гарантирован, например блик или несфокусированность на важной зоне документа сделает документ непригодным для распознавания.

В рамках проектов разрабатывался альтернативный подход, при котором используются только вычислительные мощности сопряженного с камерой устройства, что позволяет отказаться от передачи изображений. При этом возрастает автономность, поскольку распознавание можно использовать вне зависимости от наличия каналов связи. Дополнительно появляется возможность повысить качество распознавания в сложных условиях, когда ни один из кадров не может быть полностью корректно распознан *(рис. 5)*, но использование нескольких кадров видеопотока позволяет успешно решить задачу.

Основной сложностью при использовании такой схемы является существенно более скромные возможности вычислителя, поэтому обеспечение высокой скорости распознавания кадров видеопоследовательности непосредственно на устройстве при

Рис. 5. Кадры видеопоследовательности с проективно искаженным изображением документа и бликами в разных областях.

обеспечении достаточного качества распознавания является ключевым фактором успешной реализации систем распознавания документов с использованием малоразмерных цифровых камер.

Рассмотрим схему организации распознавания документа с использованием кадров видеопоследовательности. После захвата кадра проводится быстрая оценка качества изображения в целом. При неудовлетворительном качестве кадр пропускается, а параметры съемки корректируются. На достаточно качественных изображениях осуществляется детектирование четырехугольника документа. Если документ не обнаружен, то кадр пропускается. Успешно детектированная зона документа подвергается проективному исправлению. Проективно исправленная зона документа используется для выбора из набора допустимых макетов документов наиболее подходящего. Выбранный макет определяет положение зон реквизитов. Для каждой нераспознанной зоны реквизитов проводится оценка качества изображения в зоне и для достаточно качественных зон осуществляется распознавание реквизитов. Изображение зоны, содержащей еще не распознанные реквизиты, сегментируется на строки. Изображение строки, которая соответствует еще не распознанному полю, оценивается с точки зрения пригодности для дальнейшего распознавания (отсутствие засветки и бликов). Изображения строк, признанные пригодными для распознавания, сегментируются на знакоместа, которые в свою очередь распознаются. По мере появления результатов для нескольких кадров проводится межкадровое интегрирование результатов распознавания с учетом взаимного соответствия координат символов на кадрах видеопотока, языковая постобработка текстовых значений [1, 3] и проверка значения реквизитов, оценка надежности полученных результатов. В *таблице 1* приведен пример результатов, который хорошо иллюстрирует существенную зависимость финального результата от длины поля и возможностей контекстной постобработки результатов распознавания.

Таблица 1. *Оценка качества распознавания отдельных реквизитов третьей страницы паспорта гражданина РФ для 100 видеопоследовательностей*

Реквизит	Правильные распознавания, %
Дата рождения	99
Пол	99
Номер	98
Серия	97
Имя	97
Отчество	97
Фамилия	91
Место рождения	84

Важным результатом исследования стала схема использования информации о качестве изображения и его зон для активного управления устройством съемки, а также адаптивная схема с остановкой распознавания зон и отдельных реквизитов по мере получения достаточно надежного результата распознавания. Таким образом, система распознавания, ранее линейная, была трансформирована в систему с обратными связями. На *рисунке 6* представлен пример графика зависимости точности распознавания значения реквизитов в зависимости от количества обработанных кадров.

График хорошо иллюстрирует рост качества распознавания по мере увеличения количества обработанных кадров. При этом для некоторых полей требуемая точность и надежность достигается быстрее, чем для других, а своевременная остановка распознавания таких реквизитов позволяет эффективно использовать вычислительные ресурсы.

Использование методов машинного обучения

Модули распознавания на основе нейронных сетей являются наиболее эффективным инструментом для классификации изображений. Увеличение устойчивости классификаторов такого типа может достигаться путем усложнения их внутренней структуры, что неизбежно приводит к увеличению вычислительной сложности распознающего модуля и повышает требования к объему обучающей выборки.

Для достижения приемлемого времени работы на целевых устройствах в рамках проекта проведены

 DOI: 10.22204/2410-4639-2016-092-04-97-108 № 4 (92) октябрь–декабрь 2016 г.

исследования методов достижения промышленного качества распознавания при сохранении высокого быстродействия. Была показана возможность ограничения числа активных нейронов за счет использования составных классификаторов [4]. При сохранении архитектуры сети производительность может повышаться за счет понижения до 16 бит точности хранения и расчета весов (как с использованием вещественной арифметики, так и арифметики с фиксированной точкой), а также благодаря векторизации вычислений [5].

Реальные исходные данные часто содержат неточности, например для изображений отдельных символов отклонения положения охватывающего прямоугольника приводят к появлению дополнительных «полей». При неконтролируемых условиях съемки изображения принадлежащих к одному классу символов сильно варьируются, а в пространстве входных параметров классификатора наблюдается существенная неоднородность расположения данных. В то же время для многих полей документов распределение частот символов является крайне неравномерным, поэтому исходная обучающая выборка получается очень несбалансированной по числу представителей отдельных классов символов (количество представителей часто встречающихся классов может быть в несколько сот раз больше, чем количество редко встречающихся).

Собрать большую и разнообразную базу эталонных изображений документов, удостоверяющих личность, крайне сложно. Важной частью исследования было решение задачи формирования (путем синтеза) достаточно репрезентативной относительно решаемой задачи и сбалансированной внутри каждого класса и относительно различных классов выборки для обучения классификаторов на основе имеющихся данных в условиях принципиальной недостаточности исходных данных [6, 7].

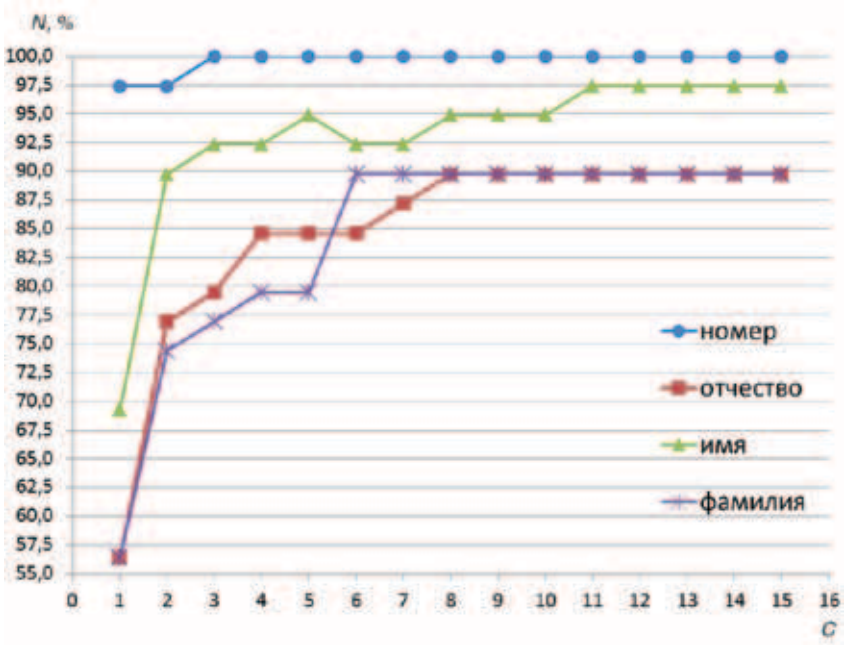


Рис. 6. Зависимость доли (N) правильно распознанных реквизитов от количества кадров (C) в последовательности.

В рамках проекта по результатам анализа искажений при съемке для моделирования изображений, схожих с реальными данными, использовались следующие методы синтеза:

1. Моделирование возникающих при съемке малоформатной камерой перспективных искажений (аффинные преобразования изображений, «сдвиг»).
2. Моделирование эффектов фотооцифровки изображения (размытие изображений с ядром Гаусса, гамма-коррекция, «стирание» локальных областей символа на изображении, случайное зашумление изображения).
3. Моделирование эффектов изменения освещения при фотосъемке (тени, изменение глобального контраста изображения).
4. Моделирование работы сегментатора (сдвиг областей полей, поворот изображения символа на случайный угол).

Синтез каждого образца для расширения обучающей выборки осуществляли путем применения к случайно выбранному изображению исходной выборки случайного поднабора из описанных выше преобразований. При этом вероятность применения каждого преобразования, а также параметры самих преобразований являлись параметрами синтеза данных и выбирались экспериментально путем минимизации ошибки классификации (на тестовых данных) при обучении на синтезированных данных. Предложенный метод синтеза данных позволяет осуществлять итерационное улучшение качества обучения классификаторов за счет дополнения обу-

чающей выборки кластерами образцов, схожих с характерными ошибками классификатора, обученного на предыдущей итерации.

Чтобы использовать классификатор в составе модуля автоматической сегментации печатных символов полей, он должен отличать символы от так называемого класса ложных сегментов, который составляют изображения фрагментов символов, пар и троек различных символов. Для надежного распознавания таких ложных сегментов необходимо присутствие в обучающей выборке достаточного числа сочетаний символов и их фрагментов. При этом исходные наборы изображений не содержат нужного числа определенных сочетаний. Репрезентативность обучающей выборки повышали генерированием искусственно «склеенных» изображений определенных символов или их фрагментов. При этом доля образцов сочетаний каждого фиксированного набора символов/фрагментов регулировалась в соответствии с частотой возникновения соответствующих ошибок классификации при тестировании на реальных данных. На *рисунке 7a* видно существенное различие фонов различных символов, что порождает видимые границы наложения в результате их «склейки» *(рис. 7b)*, значительно зашумляющие векторы входных признаков.

Для «бесшовного склеивания» изображений без образования видимых артефактов на границах наложения изображений предложен [7] метод бесшовной склейки с применением преобразования Пуассона (результат представлен на *рис. 7c*).

Модель внутренней релевантности классификатора

В системах распознавания текста для повышения точности распознавания слов существует большое количество методов, которые являются алгоритмами коррекции выходов классификатора (т.е. результатов распознавания одиночных символов), а задача коррекции, как правило, формулируется в терминах вероятностей [8]. При этом применяемые на практике методы машинного обучения дают не имеющие вероятностной природы результаты, что сильно снижает возможность использования методов статистической коррекции результатов распознавания текстовых строк.

Пусть искусственная нейронная сеть распознает K классов объектов и дает ответ в виде вектора $A=(a_1,...,a_K)$, где $a_i \in (-\infty,+\infty)$ – некоторая оценка принадлежности к классу $i \in [1,K]$. *Softmax*-преобразование в выходном слое позволяет для измеримых или заранее известных характеристик распределения, выраженных в виде положительных «весовых» значений $g_1,...,g_K$, для вектора оценок A сформировать вероятности $p_i \in (0,1)$, $\sum_{i=1}^{K} p_i = 1$, принадлежности к классам по формуле: $p_i = e^{g_i a_i} \Big/ \sum_{j=1}^{K} e^{g_j a_j}$.

Параметр внутренней релевантности τ вводится следующим образом: $p_i(\tau) = e^{g_i a_i \tau} \Big/ \sum_{j=1}^{K} e^{g_j a_j \tau}$.

Модель внутренней релевантности классификатора представляет собой использование абстрактного вещественного числа $\tau \in [0,+\infty]$, значение которого можно трактовать как количество информации, получаемое системой распознавания от классификатора. Тогда выход классификатора – это вектор-функция $p(\tau) = (p_1(\tau),...,p_K(\tau))$, причем $p_1,...,p_K$ – действительные числа в отрезке [0,1] с единичной суммой $\sum_{i=1}^{K} p_i = 1$, класс объекта определяется максимальным элементом вектора p, значение $p(0)$ соответствует вектору с нулевой информацией, поступившей от классификатора, – вектору оценок $(1/K,...,1/K)$, а значение $p(+\infty)$ соответствует вектору с максимальной информацией.

Векторы оценок, получаемые в ответе классификатора, могут быть

Рис. 7. *Склеивание фрагментов изображений: a – исходные фрагменты изображений; b – результат «наивной» склейки изображений; c – результат «бесшовной» склейки изображений [7].*

DOI: 10.22204/2410-4639-2016-092-04-97-108 № 4 (92) октябрь–декабрь 2016 г.

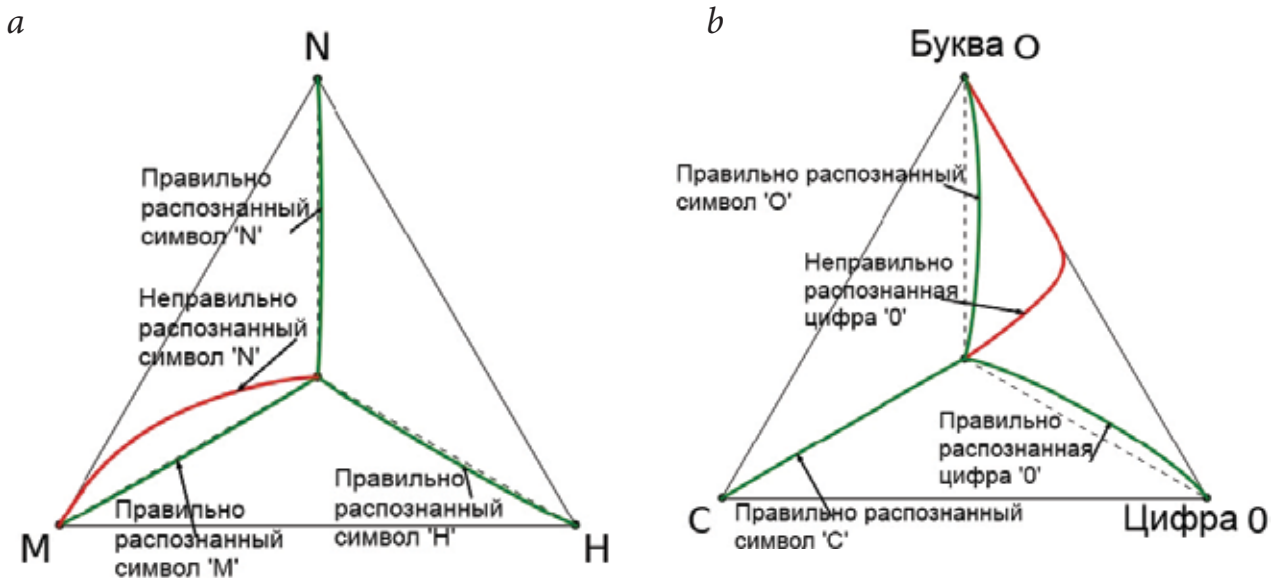


Рис. 8. *Примеры кривых, соответствующих результатам классификации: a – проекция на грань 'M'-'N'-'H'; b – проекция на грань 'C'-'0'-'O' [8].*

рассмотрены как точки в K-мерном пространстве, принадлежащие $(K\text{-}1)$-мерному симплексу с центром в точке $(1/K,..., 1/K)$ и вершинами $(1, 0,...,0)$, $(0, 1, 0...,0)$ и т.д. Вектор-функция $p(\tau)$ соответствует одномерной кривой, соединяющей центр симплекса $(\tau = 0)$ с центром одной из его $(L\text{-}1)$-мерных граней $(\tau = +\infty)$, где L – количество идентичных максимальных значений исходного вектора A, поданного на вход преобразования *softmax*. В традиционной модели выходов классификатора только одна точка на этой кривой принимается как результат классификации. Используя всю вектор-функцию, можно варьировать количество информации, полученной от метода распознавания одиночных символов.

Рассмотрим результаты работы обученной для классификации изображений символов шрифта OCR-B сверточной нейронной сети при распознавании «трудноразличимых» символов. На *рисунке 8a* представлена ортогональная проекция кривых, соответствующих полученным на выходе классификатора вектор-функциям, на двумерную грань симплекса, содержащую символы 'M', 'N' и 'H'. На *рисунке 8b* показана проекция на двумерную грань симплекса, содержащую символы 'C', 'O' (буква) и '0' (цифра).

Для правильно классифицированных изображений кривая (зеленая линия) практически не отклоняется от отрезка, соединяющего центр и вершину. При ошибочной классификации кривая (красная линия) значительно отклоняется, например на *рисунке 8a* одно из изображений символа 'N' было ошибочно классифицировано как символ 'M'. Еще более отчетливо это заметно на *рисунке 8b*: буква 'O' и цифра '0' очень трудно различимы для шрифта OCR-B. Однако кривая, соответствующая ошибочно классифицированному примеру изображения цифры '0', ярко выраженно отклоняется к «правильной» вершине, несмотря на то, что при $\tau \to +\infty$ все равно стремится к «неправильной» вершине, соответствующей букве 'O'.

Модель внутренней релевантности позволяет рассматривать выход классификатора как вектор-функцию от внутренней релевантности и может быть определена для различных алгоритмов и методов распознавания символов. С помощью такой модели можно «сдвигать» распределение оценок вероятностей принадлежности входного символа к классам при помощи варьирования значений внутренней релевантности классификатора, опираясь на известные свойства входных данных. На основе предложенной модели внутренней релевантности классификатора был разработан метод динамического варьирования релевантности, внедренный в качестве компонента в систему распознавания машиночитаемых зон с камер мобильных устройств. Метод показал хороший результат, улучшив точность системы распознавания на контрольной выборке с 99.30% до 99.67%.

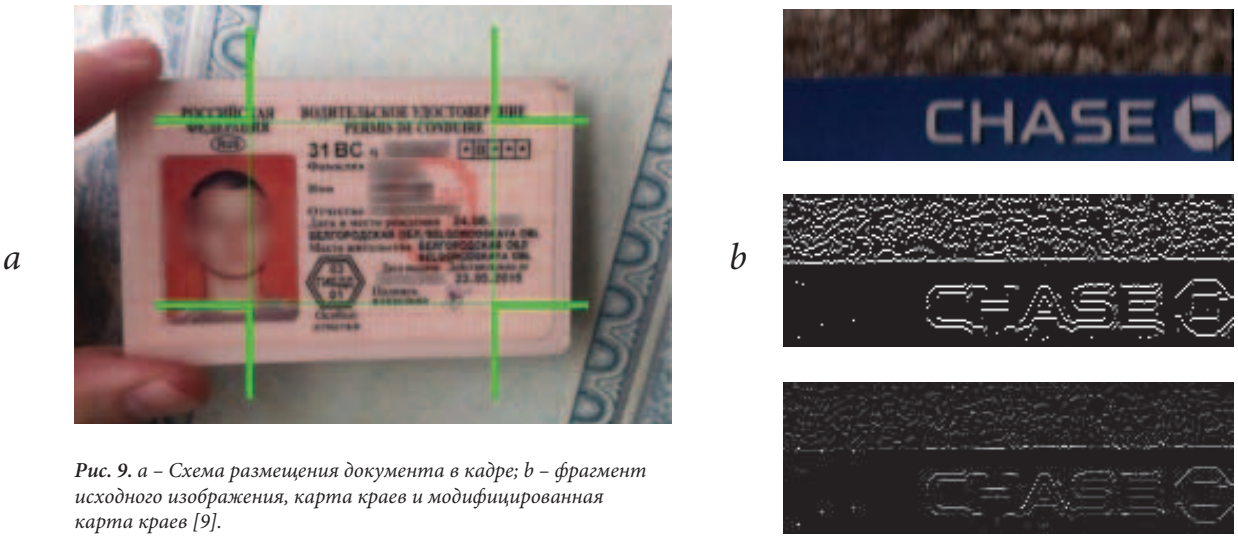


Рис. 9. *a – Схема размещения документа в кадре; b – фрагмент исходного изображения, карта краев и модифицированная карта краев [9].*

Детектирование прямоугольников на мобильных устройствах в режиме реального времени

Для достижения высоких качества и производительности [9] использовали естественные при съемке ограничения на положение документа в кадре. Документ полностью размещался на изображении и занимал его значительную часть, края документа по большей части видны *(рис. 9a)*, что обеспечивало достаточную детализацию для оптического распознавания символов с высоким качеством.

Эксперименты показали, что широкая вариативность организации сцен в неконтролируемых условиях приводит к снижению качества при использовании стандартных быстрых методов выделения границ на изображении (например детектор Канни). При помощи дополнительного статистического анализа соседних граничных областей и учета анизотропности карт градиентов при наличии преимущественной ориентации документа в кадре удалось существенно снизить влияние высокочастотной шумовой составляющей для структурированных помех *(рис. 9b)*. Для поиска фрагментов линий мы использовали модифицированное быстрое преобразование Хафа. После получения набора вертикальных и горизонтальных отрезков-кандидатов для сторон генерируется множество четырехугольников-кандидатов, для которых по допустимости проективного искажения и оценке качества выделения самих отрезков проводится выбор наиболее «похожего» на проективно искаженное изображение прямоугольника с заданным отношением сторон.

Заключение

В условиях неконтролируемых условий съемки изображения документов задача распознавания является сложной, поскольку изображения имеют высокую вариативность. В качестве основного способа достижения высокого качества авторы настоящей статьи предлагают использовать видеорежим съемки и динамическое управление ее параметрами, которое возможно при совмещении вычислителя с камерой. Распознавание непосредственно на мобильных устройствах накладывает существенные ограничения на доступные вычислительные ресурсы, однако приемлемые временные характеристики распознавания могут быть получены за счет разработанной активной схемы управления процессом обработки.

В рамках исследования предложены адаптированные к использованию видеопотока модель и методы распознавания документов, которые позволяют распознавать документы в условиях искажений и шума за счет обработки нескольких кадров видео на устройстве без передачи на сервер. Разработаны методы определения положения документа и его элементов на зашумленном изображении в режиме реального времени на мобильном устройстве, а также методы синтеза обучающих выборок для обеспечения репрезентативности и сбалансированности обучающих выборок в условиях принципиальной ограниченности объема и высокой степени несбалансированности реальных данных.

Литература

1. ***К.Б. Булатов, Д.А. Ильин, Д.В. Полевой, Ю.С. Чернышова***
Труды ИСА РАН, 2015, **65**(3), 85.
2. ***В.В. Арлазаров, А.Е. Жуковский, В.Е. Кривцов, Д.П. Николаев, Д.В. Полевой***
Информ. технол. вычисл. сист., 2014, №3, 71.
3. ***К.Б. Булатов, Д.П. Николаев, В.В. Постников***
Труды ИСА РАН, 2015, **65**(4), 68.
4. ***N. Sokolova, D.P. Nikolaev, D. Polevoy***
В *Proc. SPIE 9445 Seventh International Conference on Machine Vision (ICMV 2014), (Italy, Milan, 19–21 November, 2014)*, SPIE Publ., 2015, 944528. DOI: 10.1117/12.2180943.
5. ***E. Limonova, D. Ilin, D.P. Nikolaev***
В *Proc. SPIE 9875, Eighth International Conference on Machine Vision (ICMV 2015), (Spain, Barcelona, 19–20 November, 2015)*, SPIE Publ., 2015, 98750L. DOI: 10.1117/12.2228594.
6. ***Д.П. Николаев, Д.В. Полевой, Н.А. Тарасова***
Информ. технол. вычисл. сист., 2014, №3, 82.
7. ***А.А. Иванова, Е.Г. Кузнецова, Д.П. Николаев***
В *Сб. труд. 39-й Междисципл. шк.-конф. ИТиС 2015 «Информационные технологии и системы 2015», (РФ, Сочи, 7–11 сентября, 2015 г.)*, Москва, Изд. ИППИ им. А.А. Харкевича РАН, 2015, с. 1169–1184.
8. ***K.B. Bulatov, D.V. Polevoy***
В *Proc. ECMS 2015 29th European Conference on Modelling and Simulation (Bulgaria, Albena (Varna), 26–29 May, 2015)*, ECMS Publ., 2015, pp. 488-491. DOI: 10.7148/2015-0488.
9. ***N. Skoryukina, D.P. Nikolaev, A. Sheshkus, D. Polevoy***
В *Proc. SPIE 9445, Seventh International Conference on Machine Vision (ICMV 2014), (Italy, Milan, 19–21 November, 2014)*, SPIE Publ., 2015, 94452A. DOI: 10.1117/12.2181377.

English

Key Aspects of Document Recognition Using Small Digital Cameras*

Dmitriy V. Polevoy –
National University of Science and Technology MISIS
4, Leninskiy Ave.,
Moscow, 119049, Russia
e-mail: dvpsun@gmail.com

Konstantin B. Bulatov –
National University of Science and Technology MISIS
4, Leninskiy Ave.,
Moscow, 119049, Russia
e-mail: hpbuko@gmail.com

Natalya S. Skoryukina –
National University of Science and Technology MISIS
4, Leninskiy Ave.,
Moscow, 119049, Russia
e-mail: skleppy.inc@gmail.com

Timofey S. Chernov –
National University of Science and Technology MISIS
4, Leninskiy Ave.,
Moscow, 119049, Russia
e-mail: chernov.tim@gmail.com

Vladimir V. Arlazarov –
Moscow Institute of Physics and Technology (State University)
9, Institutskiy Per., Dolgoprudny,
Moscow Region, 141700, Russia
e-mail: vva777@gmail.com

Aleksander V. Sheshkus –
Smart Engines Service Ltd.
9, 60-letiya Octyabrya Ave.,
Moscow, 117312, Russia
e-mail: asheshkus@smartengines.biz

Abstract

The paper summarizes a series of studies solving the recognition problem of identity documents images captured by small digital cameras. The main features of the documents and the shooting process and their impact on the recognition of the image are shown. The authors offer a new approach to building the recognition system as a feedback system.

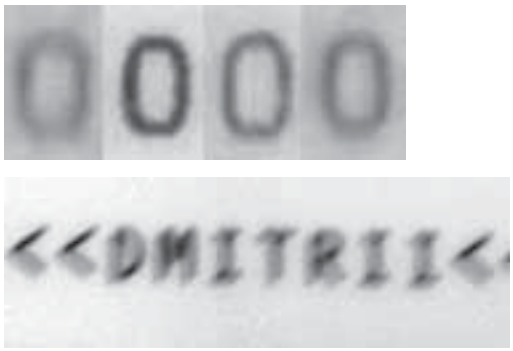
Keywords: document recognition, on-device optical character recognition, inter-frame integration.

* *The work was financially supported by RFBR (projects 13-07-12171, 13-07-12172 and 14-07-00730).*

Images & Tables

Fig. 1. Fragments of document images with different types of backgrounds and fonts.

a



b

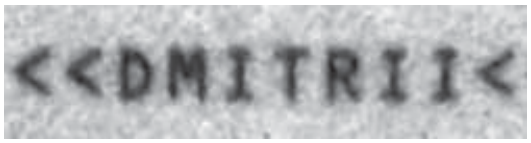


Fig. 2. Fragments of images of documents: a – with blur; b – with the example of digital noise.

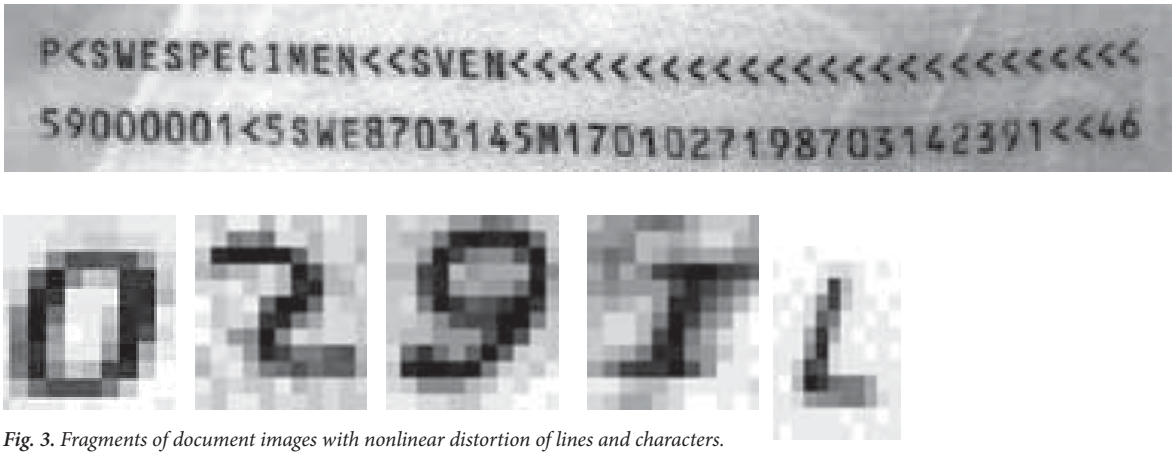


Fig. 3. Fragments of document images with nonlinear distortion of lines and characters.

Fig. 4. Fragments of images of protected documents.

DOI: 10.22204/2410-4639-2016-092-04-97-108 № 4 (92) октябрь–декабрь 2016 г.

Fig. 5. Video sequence frames with projectively distorted images of the document and flash glares in different areas.

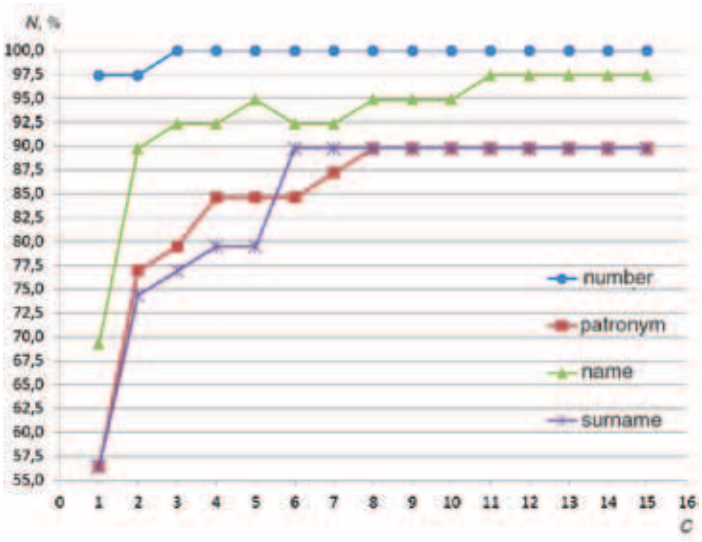


Fig. 6. The proportion (N) of correctly recognized documents' requisites depending on the number of frames (C) in the sequence.

Fig. 7. Bonding of the images' fragments: a – original image slices; b – the result of a "naïve" gluing images; c – the result of the "seamless" gluing of the images [7].

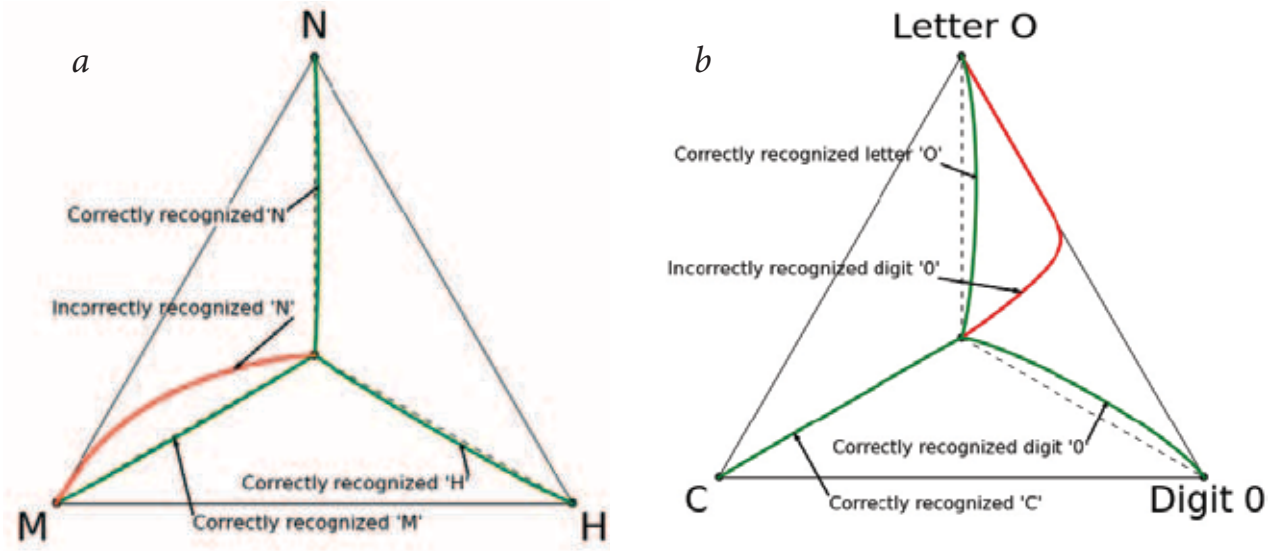


Fig. 8. Examples of curves corresponding to the classification results: a – projection on the 'M' - 'N' - 'H' facet; b – projection on the 'C' - '0' - 'O' facet [8].

a

b

Fig. 9. a – The positioning of the document in the frame; b – a fragment of the original image, the edges of the card and a modified edge map [9].

***Table 1.** Quality of the Russian Federation citizen's passport fields recognition through the example of 100 video sequences*

Passport field	Proportion of correct recognition, %
Birthdate	99
Gender	99
Number	98
Series	97
Name	97
Patronym	97
Surname	91
Birthplace	84

References

1. ***K.B. Bulatov, D.V. Polevoy, D.A. Ilin, Y.S. Chernyshova***
Proc. ISA RAS [Trudy ISA RAN], 2015, **65**(3), 85 (in Russian).
2. ***V.V. Arlazarov, A.E. Zhukovsky, V.E. Krivtsov, D.P. Nikolaev, D.V. Polevoy***
J. Information Technology and Computer Systems [Informatsionnye tekhnologii i vychislitelnye sistemy], 2014, №3, 71 (in Russian).
3. ***K.B. Bulatov, D.P. Nikolaev, V.V. Postnikov***
Proc. ISA RAS [Trudy ISA RAN], 2015, **65**(4), 68 (in Russian).
4. ***N. Sokolova, D.P. Nikolaev, D. Polevoy***
In *Proc. SPIE 9445 Seventh International Conference on Machine Vision (ICMV 2014), (Italy, Milan, 19–21 November, 2014),* SPIE Publ., 2015, 944528. DOI: 10.1117/12.2180943.
5. ***E. Limonova, D. Ilin, D.P. Nikolaev***
In *Proc. SPIE 9875, Eighth International Conference on Machine Vision (ICMV 2015), (Spain, Barcelona, 19–20 November, 2015),* SPIE Publ., 2015, 98750L. DOI: 10.1117/12.2228594.
6. ***D.P. Nikolaev, D.V. Polevoy, N.A. Tarasova***
J. Information Technology and Computer Systems [Informatsionnye tekhnologii i vychislitelnye sistemy], 2014, №3, 82 (in Russian).
7. ***A.A. Ivanova, E.G. Kuznetsova, D.P. Nikolaev***
In *Proc. 39th School-Conf. IT&S 2015 "Information Technologies and Systems 2015" [Informatsionnye tekhnologii i sistemy], (RF, Sochi, 7–11 September, 2015),* RF, Moscow, A.A. Kharkevich IITP RAS Publ., 2015, pp. 1169-1184 (in Russian).
8. ***K.B. Bulatov, D.V. Polevoy***
In *Proc. ECMS 2015 29th European Conference on Modelling and Simulation (Bulgaria, Albena (Varna), 26–29 May, 2015),* ECMS Publ., 2015, pp. 488-491. DOI: 10.7148/2015-0488.
9. ***N. Skoryukina, D.P. Nikolaev, A. Sheshkus, D. Polevoy***
In *Proc. SPIE 9445, Seventh International Conference on Machine Vision (ICMV 2014), (Italy, Milan, 19–21 November, 2014),* SPIE Publ., 2015, 94452A. DOI: 10.1117/12.2181377.

Методы интеграции результатов распознавания текстовых полей документов в видеопотоке мобильного устройства *

К.Б. Булатов, В.Ю. Кирсанов, В.В. Арлазаров, Д.П. Николаев, Д.В. Полевой

Одной из ключевых подсистем в системах распознавания документов в видеопотоке является подсистема интеграции результатов распознавания на одиночных изображениях. В работе рассматривается задача интеграции результатов распознавания текстовых полей в видеопотоке, полученном при помощи камеры мобильного устройства, описываются простейшие алгоритмы интеграции и предлагается алгоритм, основанный на выравнивании входных строк при помощи модифицированного редакционного расстояния. Приводится сравнительный анализ результатов работы алгоритмов на примере задачи распознавания полей паспорта гражданина Российской Федерации в видеопотоке мобильного устройства.

Ключевые слова: оптическое распознавание символов, анализ документов, распознавание в видеопотоке, анализ видеопотока, мобильные устройства.

** Работа выполнена при финансовой поддержке РФФИ (проекты №№ 13-07-12170, 13-07-12171 и 13-07-12172).*

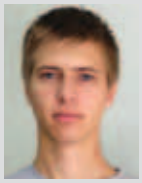
Введение

В 2000-х гг. появляется широкий интерес к методам автоматического ввода документов с использованием мобильных устройств. Это обусловлено быстро растущими вычислительными возможностями таких широко распространенных мобильных устройств, как смартфоны и портативные планшетные компьютеры, а также увеличивающимися техническими возможностями цифровых камер, установленных на этих устройствах. Интерес к системам электронного документооборота, в частности, к методам автоматического ввода документов, применительно к мобильным устройствам также связан с развитием систем распространения мобильных приложений – как корпоративных, так и нацеленных на широкую публику.

В корпоративном секторе повышается интерес к реализации делопроизводства (или его части) на основе мобильного документооборота – разновидности электронного документооборота, пользователи которого получают возможность проводить операции с электронными документами при помощи различных мобильных устройств. Естественно, встает задача реализации систем автоматического ввода документов, использующих цифровые камеры мобильных устройств в качестве «сканирующего» устройства, – оцифровка документа проводится путем видео- или фотосъемки оригинала.

Среди обычных пользователей таких мобильных устройств, как смартфоны или планшетные компьютеры, возрастает интерес к приобретению товаров и услуг посредством транзакций через интернет-сервисы, доступные для персональных мо-

БУЛАТОВ
Константин Булатович
Национальный исследовательский технологический университет МИСиС



КИРСАНОВ
Владимир Юрьевич
Московский физико-технический институт

АРЛАЗАРОВ
Владимир Викторович
Московский физико-технический институт

НИКОЛАЕВ
Дмитрий Петрович
Институт проблем передачи информации им. А.А. Харкевича РАН

ПОЛЕВОЙ
Дмитрий Валерьевич
Национальный исследовательский технологический университет МИСиС

бильных устройств. В большинстве случаев заключение таких сделок подразумевает ввод реквизитов ряда документов (к примеру документа, удостоверяющего личность, банковской карты и т.д.), причем ввод этих данных зачастую требуется проводить неоднократно, так как хранение этих данных в памяти мобильного устройства может привести к их утечке и использованию злоумышленниками. Хранение «чувствительных» персональных данных на интернет-серверах строго ограничивается законодательством и так же, хоть и в меньшей степени, чем на локальном устройстве, подвержено атакам со стороны мошенников. Это приводит к тому, что методы автоматического ввода документов, ориентированные на мобильные устройства, приобретают актуальность не только в корпоративной сфере, но и в сфере массовой электронной коммерции.

Подавляющее большинство работ, связанных с автоматическим вводом и распознаванием документов на мобильных устройствах, рассматривает фотографию документа как его электронное представление и отмечает [1, 2] трудности, связанные с подготовкой образа документа к распознаванию и с самим распознаванием.

Изображения документов, сделанные камерой мобильного устройства, гораздо более низкого качества, чем изображения, получаемые на традиционном цифровом сканере. В случае мобильных устройств на этапе подготовки изображения к распознаванию приходится сталкиваться с такими проблемами, как неравномерное освещение сцены, проективные искажения документа, нелинейные искажения документа (например изгибы бумажного носителя), искажения, обусловленные движением камеры, зашумление, дефокусировка, световые блики и т.п. *(рис. 1)*. Все это приводит к тому, что традиционные методы предварительной обработки изображения, применяемые в системах автоматического ввода документов с использованием цифровых сканеров, не дают необходимого эффекта, и появляется необходимость в специальных методах, позволяющих увеличить точность и надежность распознавания.

Одним из способов нивелировать эффекты, создаваемые описанными проблемами, является рассмотрение цельного видеопотока в качестве цифрового образа документа. Полагая, что на разных кадрах одного видеопотока присутствует графический образ одного и того же документа (или фрагмента документа), возможно, с различными искажениями и проводя оптическое распознавание документа и его полей на нескольких кадрах, мы можем улучшить точность и надежность распознавания, интегрируя результаты. Под «интеграцией» в данном контексте будем понимать восстановление единого результата распознавания объекта в видеопотоке из частных результатов распознавания этого же объекта на отдельных кадрах видеопотока.

В данной работе будет описан один из алгоритмов интеграции результатов распознавания текстовых полей в видеопотоке и представлены результаты его использования.

Простейшие алгоритмы интеграции

Пусть для каждого кадра f входного видеопотока задан результат распознавания текстового поля (модель гипотез) в виде последовательности ячеек $s^f = C_1^f,...,C_N^f$, а каждая ячейка соответствует результату распознавания некоторого символа и представляет собой множество альтернатив и их оценок:

$$C_i^f = \left\{ \left\langle a_{i_1}, q_{i_1}^f \right\rangle,...,\left\langle a_{i_K}, q_{i_K}^f \right\rangle \right\},$$

где a_{i_j} – j-й символ алфавита $\sum$, $q_{i_j}^f$ – оценка этого символа. Для удобства

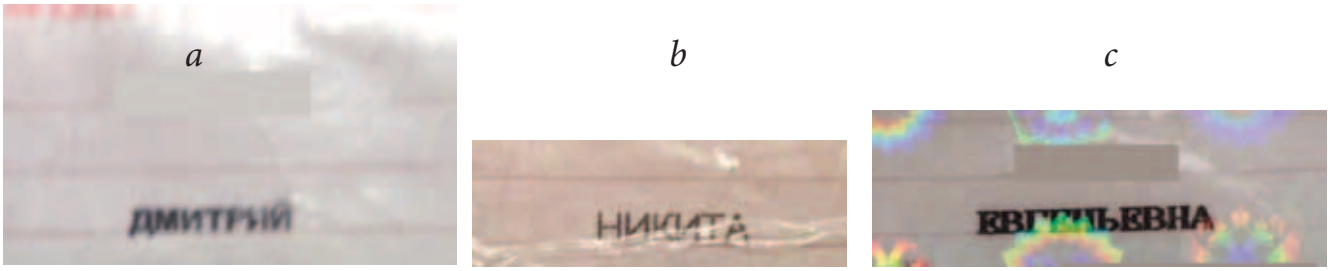


Рис. 1. *Примеры дефектов изображений, полученных при помощи камеры мобильного устройства: a – дефокусировка и блики, b – дефекты и блики, c – голографическая защита.*

DOI: 10.22204/2410-4639-2016-092-04-109-115 № 4 (92) октябрь–декабрь 2016 г.

будем считать, что каждая ячейка содержит K альтернатив и все оценки альтернатив принимают положительное значение. Под оценкой (весом) строки будем понимать произведение оценок каждого из символов этой строки.

Среди простейших методов интеграции можно выделить следующие три: выбор наилучшего результата, посимвольная интеграция, мультимодальная интеграция.

Выбор наилучшего результата. Интеграция методом выбора наилучшего результата является простейшим способом выделения единственного результата из последовательности гипотез, вычисленных на каждом кадре. При условии, что определен некоторый вычислимый количественный критерий, который можно считать устойчивым предиктором качества распознавания, результатом интеграции является гипотеза, обладающая максимальным значением этого предиктора.

Пусть задана некоторая функция $conf(s^f)$, аргументом которой является результат s^f распознавания текстового поля, а значением – вещественное число (оценка) в отрезке [0,1]. Функцию $conf(s^f)$ будем считать устойчивым предиктором качества распознавания, если функция $P_{conf}(x) = P(s^f|conf(s^f) \geq x)$ – оценка вероятности правильного распознавания строки при условии, что ее оценка больше некоторого числа x, монотонно возрастает (x также является вещественным числом в отрезке [0.1]). Достаточно устойчивыми примерами таких дескрипторов являются средняя оценка старшей альтернативы по строке $conf(s^f) = \frac{1}{N}\sum_{i=1}^{N} \max_j q_{i_j}^f$, а также средняя разница между двумя старшими альтернативами ячеек строки:

$$conf(s^f) = \frac{1}{N}\sum_{i=1}^{N} \times$$

$$\times\left[\max_j q_{i_j}^f - \max_{j \neq \arg\max_k q_{i_j}^f} q_{i_j}^f\right].$$

Посимвольная интеграция. Следующим простейшим методом интеграции результатов распознавания является посимвольная интеграция распознанных строк. Пусть заданы два входных результата распознавания s^f и s^g одного и того же текстового поля с двух различных кадров f и g видеопотока $s^f = C_1^f,..., C_N^f$, $s^g = C_1^g,..., C_N^g$ одинаковой длины N. При посимвольной интеграции формируется такая последовательность ячеек $C_1,..., C_N$, что каждая ячейка C_i образуется только с использованием соответствующих ячеек исходных последовательностей C_1^f и C_1^g путем усреднения оценок соответствующих символов:

$$C_i = \left\{\left\langle a_{i_1}, \frac{1}{2}(q_{i_1}^f + q_{i_1}^g)\right\rangle, ..., \left\langle a_{i_K}, \frac{1}{2}(q_{i_K}^f + q_{i_K}^g)\right\rangle\right\}.$$

Мультимодальная интеграция. Очевидным недостатком простой посимвольной интеграции является то, что такой алгоритм не сможет работать с результатами распознавания одного и того же текстового поля разных кадров, неодинаковых по длине. Результаты распознавания могут иметь разную длину для полей с немоноширинной печатью и неизвестным заранее шаблоном ввиду ошибок на этапе сегментации изображения поля на отдельные символы.

Для того чтобы интегрировать результаты распознавания текстовых полей в видеопотоке с учетом возможных ошибок сегментации, предлагается использовать метод мультимодальной интеграции.

Изначально результаты распознавания с различных кадров видеопотока кластеризуются по некоторому заранее определенному критерию, позволяющему использовать метод простой посимвольной интеграции внутри каждого кластера. Простейшим критерием является длина распознанной строки (количество ячеек в результате распознавания), однако критерий можно модифицировать для увеличения устойчивости. После этого внутри каждого кластера проводится простая посимвольная интеграция строк (уже одинаковой длины). Далее для получения финального результата интеграции результаты кластеров интегрируются методом выбора результата с максимальным значением предиктора качества $conf(s)$.

Развитие алгоритмов интеграции

Основной проблемой мультимодального метода интеграции является то, что различные ошибки сегментации текстового поля на двух разных кадрах видеопотока могут привести к смещению друг относительно друга цепочек символов из двух последовательностей, помещенных в один кластер, и при

внутрикластерной интеграции совместно будут рассматриваться результаты распознавания, соответствующие различным символам.

Для того чтобы избежать этой проблемы, предлагается предварительно выравнивать полученные строки, достигая минимального «редакционнго расстояния» между строками-результатами, и затем интегрировать выровненные строки посимвольным методом. Под редакционным расстоянием понимается минимальная суммарная стоимость операций, которые необходимо совершить для приведения одной строки к другой. Операции всего три: добавление символа, удаление символа и его замена на другой символ алфавита. Стоимости всех операций принимаются за единицу.

Похожий подход был использован в работе [3] в качестве метода интеграции результатов работы нескольких распознающих алгоритмов в задаче распознавания речи. В рамках используемого подхода все результирующие строки объединялись в единую сеть словесных переходов (Word Transition Network, WTN), в которой строки выравнивались относительно первой обработанной путем применения алгоритма динамического программирования для поиска расстояния Левенштейна. Далее, при помощи механизма простого голосования, по WTN строилась результирующая строка. Результат данного алгоритма является неустойчивым к порядку добавления интегрируемых строк (так как на каждой итерации очередная строка выравнивается с результатом интегрирования предыдущих). Идея получила развитие в работе [4], в которой авторы использовали подобный алгоритм в системе, распознающей текст на изображениях. Важным отличием стало то, что в их работе вместо интеграции простых символьных строк (с которыми работали авторы работы [3]) проводилась интеграция результатов работы разных алгоритмов распознавания, представляющих собой строчки с альтернативами для каждого символа (подобно модели результата распознавания текстовой строк в главе «Простейшие методы интеграции»). Однако на этапе выравнивания строк при помощи алгоритма динамического программирования учитывалось редакционное расстояние между строками, образованными символами, соответствующими лишь максимальным альтернативам.

Для того чтобы выравнивать результаты распознавания полей с разных кадров по длине, расширим алфавит специальным «пустым» символом. Далее при помощи алгоритма динамического программирования поиска редакционного расстояния проводится вставка ячейки с пустым символом в несколько мест обеих последовательностей ячеек. При этом минимизируется редакционное расстояние между финальными последовательностями ячеек и попарные редакционные расстояния между любыми префиксами этих последовательностей.

При расчете редакционного расстояния для выравнивания строк в качестве стоимости вставки «пустого» символа можно принять заранее определенный настраиваемый параметр $E \in [0.1]$. В качестве расстояния (меры различия) между ячейками предлагается использовать полусумму модулей разности оценок соответствующих альтернатив:

$$\rho(C_i, C_j) = \frac{1}{2}\sum_{m=1}^{K}\left|q_{i_m} - q_{j_m}\right|.$$

После процедуры выравнивания полученные последовательности ячеек имеют равную длину и интегрируются методом простого посимвольного интегрирования (см. главу «Простейшие методы интеграции»).

Сравнительный анализ качества работы предложенного метода показал, что, в отличие от мультимодального подхода, данный метод учитывает результаты распознавания поля на всех кадрах, а не только внутри «усеченного» кластера. Выравнивание строк перед посимвольной интеграцией позволяет не только учесть результаты распознавания некоторых символов, которые не были учтены в выбранном кластере мультимодальной интеграции из-за сегментационной ошибки в других частях поля, а также избежать интеграции ячеек, соответствующих различным символам, происходящей из-за ошибок на этапе кластеризации.

Результаты сравнения качества работы двух методов представлены в *таблице 1*. В анализе использовался набор данных из 235 видеозаписей съемки различных паспортов гражданина Российской Федерации, сделанных при помощи камеры мобильного устройства. В каждой видеозаписи проводилось распоз-

DOI: 10.22204/2410-4639-2016-092-04-109-115 № 4 (92) октябрь–декабрь 2016 г.

нование полей «Имя», «Фамилия», «Отчество» и «Место рождения» независимо на 50 кадрах. Ввиду нерегулярной печати, использования различных шрифтовых гарнитур, а также дефектов съемки при помощи камеры мобильного устройства при интеграции результатов распознавания этих полей должны учитываться не только возможные ошибки распознавания одиночных символов, но и возможные ошибки сегментации изображения на отдельные символы.

Сравнительный анализ, приведенный в *таблице 1*, показывает преимущество алгоритма интеграции, основанного на выравнивании строк при помощи модифицированного редакционного расстояния, для трех испытуемых полей документа («Имя», «Фамилия», «Место рождения»). Преимущество мультимодального подхода для поля «Отчество», вероятно, связано со специфической морфологической структурой отчеств в русском языке и требует отдельного исследования.

Таблица 1. *Сравнение результатов работы алгоритмов мультимодальной интеграции и интеграции на основе выравнивания строк при помощи модифицированного редакционного расстояния*

Поле	Точность распознавания поля, %	
	Мультимодальная интеграция	Модифицированный алгоритм
Имя	92.34	93.19
Фамилия	89.36	89.79
Отчество	91.91	90.21
Место рождения	45.53	48.09

Заключение

В работе была рассмотрена проблема интеграции результатов распознавания текстовых полей в видеопотоке, полученном при помощи камеры мобильного устройства, а также несколько алгоритмов интеграции и их особенности, предложен алгоритм на основе выравнивания входных последовательностей результатов распознавания символов при помощи модифицированного редакционного расстояния, представлено его обоснование. Сравнительный анализ методов интеграции показал улучшение точности распознавания полей паспорта гражданина Российской Федерации в видеопотоке на мобильном устройстве при использовании предложенного алгоритма (по сравнению с более простыми методами) для трех из четырех испытуемых полей паспорта РФ («Имя», «Фамилия» и «Место рождения»), из чего можно сделать вывод о целесообразности использования предложенного алгоритма в промышленных системах распознавания текстовых полей в видеопотоке.

В рамках дальнейших исследований предполагается разработать комбинированный метод на основе первоначальной слабой кластеризации входных последовательностей с последующей внутрикластерной интеграцией предложенным в данной работе методом. Кроме того, планируется модифицировать метод вычисления расстояния между ячейками выравниваемых последовательностей путем использования информации о геометрическом расположении найденных ячеек символов с нормализацией в рамках единой системы координат документа в видеопотоке.

Литература

1. ***E.D. Haritaoglu, I. Haritaoglu***
В *Proc. 2005 Symp. Document Image Understanding Technology (SDIUT05), (USA, Maryland, Adelphi, 2–4 November, 2005),* USA, University of Maryland Publ., 2005, pp. 55–62.
2. ***M. Hsueh***
Interactive Text Recognition and Translation on a Mobile Device, EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2011-57, 2011, 13 pp. (http://www2.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-57.html).
3. ***J.G. Fiscus***
A Post-Processing System To Yield Reduced Word Error Rates: Recognizer Output Voting Error Reduction (ROVER), NIST Information Technology Laboratory, 1997, 8 pp. (http://citeseer.ist.psu.edu/viewdoc/download?doi=10.1.1.23.5624&rep=rep1&type=pdf). DOI: 10.1.1.23.5624.
4. ***R. Prasad, S. Saleem, E. MacRostie, P. Natarajan, M. Decerbo***
В *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing ICASSP–2008 (USA, Nevada, Las Vegas, 30 March–4 April, 2008),* IEEE Publ., 2008, pp. 1357–1360. DOI: 10.1109/ICASSP.2008.4517870.

English

Methods for Integration the Results of the Documents Text Fields Recognition in the Videostream of a Mobile Device*

Konstantin B. Bulatov –
National University of Science and Technology MISIS
4, Leninskiy Ave., Moscow, 119049, Russia
e-mail: hpbuko@gmail.com

Vladimir Yu. Kirsanov –
Moscow Institute of Physics and Technology
9, Institutskiy Per., Dolgoprudny,
Moscow Region, 141700, Russia
e-mail: kirsanovvl@gmail.com

Vladimir V. Arlazarov –
Moscow Institute of Physics and Technology
9, Institutskiy Per., Dolgoprudny,
Moscow Region, 141700, Russia
e-mail: vva777@gmail.com

Dmitriy P. Nikolaev –
A.A. Kharkevich Institute
for Information Transmission
Problems RAS
19-1, Bolshoy Karetny Per.,
Moscow, 127051, Russia
e-mail: d.p.nikolaev@gmail.com

Dmitriy V. Polevoy –
National University of Science
and Technology MISIS
4, Leninskiy Ave., Moscow, 119049, Russia
e-mail: dvpsun@gmail.com

Abstract

One of the key subsystems in optical character recognition systems processing videostreams is a subsystem that integrates the results of single-frame recognition. In the paper authors considered the problem of integration of text fields identification results in the videostream received *via* a mobile device camera, described the simplest integration algorithms, and offered the algorithm based on the alignment of the input strings by means of a modified edit distance. A comparative analysis of the algorithms' outputs (the task of Russian Federation citizen passport text fields recognition in the videostream received *via* a mobile device camera was considered as an example) was performed.

Keywords: optical character recognition, document analysis, recognition in videostream, videostream analysis, mobile devices.

* *The work was financially supported by RFBR (projects 13-07-12170, 13-07-12171 and 13-07-12172).*

Images & Tables

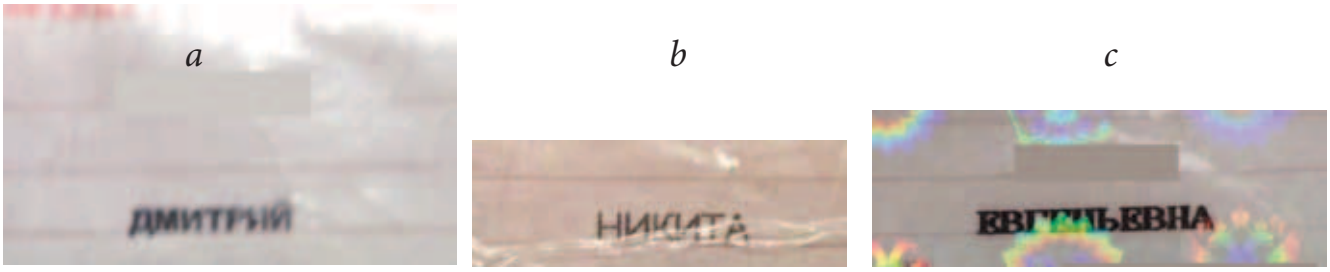


Fig. 1. Examples of defects of images captured with a mobile device's camera: a – defocus and flares, b – defects and flares, c – holographic security elements obstruction.

Table 1. *Comparison of performance of the multimodal integration algorithm with the integration algorithm based on the alignment of the input strings by means of a modified edit distance*

Document field	Precision of document text field recognition, %	
	Multimodal integration	Modified algorithm
First name	92.34	93.19
Last name	89.36	89.79
Patronym	91.91	90.21
Birthplace	45.53	48.09

References

1. E.D. Haritaoglu, I. Haritaoglu
In *Proc. 2005 Symp. Document Image Understanding Technology (SDIUT05), (USA, Maryland, Adelphi, 2–4 November, 2005),* USA, University of Maryland Publ., 2005, pp. 55–62.

2. M. Hsueh
Interactive Text Recognition and Translation on a Mobile Device, EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2011-57, 2011, 13 pp. (http://www2.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-57.html).

3. J.G. Fiscus
A Post-Processing System To Yield Reduced Word Error Rates: Recognizer Output Voting Error Reduction (ROVER), NIST Information Technology Laboratory, 1997, 8 pp. (http://citeseer.ist.psu.edu/viewdoc/download?doi=10.1.1.23.5624&rep=rep1&type=pdf). DOI: 10.1.1.23.5624.

4. R. Prasad, S. Saleem, E. MacRostie, P. Natarajan, M. Decerbo
In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing ICASSP–2008 (USA, Nevada, Las Vegas, 30 March–4 April, 2008),* IEEE Publ., 2008, pp. 1357–1360. DOI: 10.1109/ICASSP.2008.4517870.

Алгоритм применения *N*-грамм для корректировки результатов распознавания *

Т.В. Манжиков, О.А. Славин, И.А. Фараджев, И.М. Янишевский

В работе исследуется применение *N*-грамм для корректировки результатов распознавания образов слов документов на примере полей паспорта гражданина РФ. Для триграмм приводятся два алгоритма корректировки результатов распознавания. Один из них базируется на использовании вероятностей триграмм в сочетании с оценками распознавания, также интерпретируемыми как вероятности. Второй алгоритм основан на определении маргинальных распределений и вычислениях на графах на основе байесовских сетей. Приводятся результаты экспериментов применения алгоритмов и сравнение характеристик обоих алгоритмов.

Ключевые слова: распознавание символов, *N*-грамма, триграмма, маргинальное распределение, вычисления на графах.

** Работа выполнена при финансовой поддержке РФФИ (проекты №№ 13-07-12170, 13-07-12171 и 16-07-01051).*

Введение

Механизм использования *N*-грамм в алгоритмах анализа текста является хорошо изученным в настоящее время. Формально *N*-граммой на алфавите *A*, состоящем из конечного числа символов, называется строка из *N* символов, принадлежащих алфавиту *A*.

В настоящей статье предполагается возможность априори применения *N*-грамм для поиска ошибок в словах, основанная на эффективности такого применения при компьютерной проверке слов программами типа спелл-чекеров (spelling checker) [1, 2].

Мы будем рассматривать только *N*-граммы, состоящие из букв русского алфавита. Другие виды *N*-грамм, базирующиеся на словах естественного языка или фонемах речи, не рассматриваются, поскольку предметом статьи является исследование возможностей применения *N*-грамм для постобработки результатов распознавания. Рассматриваются только триграммы, для других видов *N*-грамм возможны аналогичные обобщения.

Применение *N*-граммы при проверке результатов распознавания

Рассмотрим представление результатов распознавания слова из n букв в виде нескольких альтернатив (c_j, p_j):

$$(c_1, p_1), (c_2, p_2), \ldots, (c_n, p_n), \qquad (1)$$

где c_j – код символа, а p_j – оценка надежности распознавания символа (далее – оценки распознавания). При этом $c_j \in A_r$, где A_r – алфавит распознавания, а все оценки распознавания обладают нормировкой ($0 \le p_j \le 1$) и монотонностью (если $p_i < p_j$, то символ c_i считается распознанным надежнее, чем символ c_j).

Предполагается наличие некоторого механизма распознавания, который в образах символов и слов осуществляет сегментацию границ символов, извлечение символов и их распознавание. Полагаем, что механизм распознавания может ошибаться, но ошибки распознавания могут успешно корректироваться во время постобработки аналогично словарным механизмам, описанным в работах [3, 4].

Модель триграмм базируется на наборах $G_i = \{g_{i1}, g_{i2}, g_{i3}, \theta(g_{i1}, g_{i2}, g_{i3})\}$,

МАНЖИКОВ
Тэмуджин Валерьевич
Национальный исследовательский технологический университет МИСиС

СЛАВИН
Олег Анатольевич
Институт системного анализа ФИЦ «Информатика и управление» РАН

ФАРАДЖЕВ
Игорь Александрович
Институт системного анализа ФИЦ «Информатика и управление» РАН

ЯНИШЕВСКИЙ
Игорь Михайлович
Институт системного анализа ФИЦ «Информатика и управление» РАН

$g_{i1} \in A_g$, $g_{i2} \in A_g$, $g_{i3} \in A_g$, где A_g – словарный алфавит, g_i – символ, а $\theta(g_{i1}, g_{i2}, g_{i3})$ – вес триграммы. Совокупность всех триграмм назовем словарем триграмм $G = \{G_1, G_2, \ldots\}$. Триграммы могут быть упорядочены по весу, причем больший вес означает большую частоту встречаемости триграммы в естественном языке, точнее, в некотором корпусе текстов на естественном языке.

Рассмотрим задачу применения триграмм для поиска и исправления одиночных ошибок распознавания, полагая, что у ошибочного распознанного символа соседние с ним символы распознаны без ошибок.

Общеизвестным является следующий способ поиска ошибок в слове $c_1, c_2, \ldots, c_n$ с помощью словаря триграмм: исходя из предположения, что алфавиты A_g и A_r совпадают, каждую тройку символов c_{k-1}, c_k, c_{k+1} $(1 < k < n)$ сравним с каждой из триграмм $G_i \in G$ и при различии c_k и g_i зафиксируем ошибку. Учитывая возможность соотнесения каждого символа с несколькими триграммами (от одной до трех), зафиксируем ошибку в случае его непопадания ни в одну из возможных триграмм. Способ исправления одиночной ошибки состоит в замене ошибочного символа c_k на соответствующий символ из триграммы (поскольку подходящих триграмм может быть несколько, выберем триграмму с наибольшим весом).

Для результатов распознавания с альтернативами возможно следующее обобщение описанного способа исправления ошибок: после того как найден ошибочный символ, выбираем с помощью словаря триграмм символ одной из альтернатив.

Эвристический алгоритм корректировки результатов распознавания с помощью триграмм

Рассмотрим способ применения словарей триграмм для корректировки результатов распознавания полей структурированных документов. В качестве примера рассмотрим паспорт гражданина РФ, в котором содержатся поля «Фамилия», «Имя», «Отчество».

Сформируем словари триграмм следующим способом. Зададим перечень слов, составляющих тематический словарь V (например словарь фамилий). Для каждого слова $W = \{c_1, c_2, \ldots, c_n\}$ в котором $c_i \in A$, A – алфавит, содержащий возможные символы V, для рассматриваемых словарей это буквы русского языка и, возможно, символ «–», добавим два фиктивных символа c_F в начале и в конце слов: $W' = \{c_F, c_F, c_1, c_2, \ldots, c_n, c_F, c_F\}$. Это необходимо для того, чтобы каждому из символов c_i можно было сопоставить триграммы $<c_{i-2}, c_{i-1}, c_i>$ и $<c_i, c_{i+1}, c_{i+2}>$, например c_1 можно сопоставить триграмму $<c_F, c_F, c_1>$. Далее будем рассматривать слова с добавленными в алфавит A символами c_F.

Подсчитаем для каждой тройки символов $<g_1, g_2, g_3>$ $(g_1 \in A, g_2 \in A, g_3 \in A)$ частоту встречаемости этой тройки во всех словах W'. Получим словарь триграмм, состоящий из $\{g_{i1}, g_{i2}, g_{i3}, \theta(g_{i1}, g_{i2}, g_{i3})\}$ с нормированными весами (потенциалами) $\theta(g_{i1}, g_{i2}, g_{i3})$, сумма которых равна единице. Очевидно, что $0 \le \theta(g_{i1}, g_{i2}, g_{i3}) \le 1$.

Рассмотрим результаты распознавания слова в виде набора $\{(c^1_j, p^1_j), (c^2_j, p^2_j), \ldots, (c^q_j, p^q_j)\}$, где q является количеством допустимых альтернатив распознавания, а оценки p^k_j могут быть интерпретированы как вероятности, в том смысле, что выполнены условия:

$$0 \le p^k_j \le 1,$$

$$\sum_{k=1}^{q} p^k_j = 1, = \overline{1, n}.$$

Тогда слово из n символов, каждый из которых имеет q альтернатив, можно представить следующим образом:

$$\begin{gathered}\{(c^1_1, p^1_1), (c^1_2, p^1_2), \ldots, (c^1_n, p^1_n)\},\\ \{(c^2_1, p^2_1), (c^2_2, p^2_2), \ldots, (c^2_n, p^2_n)\},\\ \ldots\\ \{(c^q_1, p^q_1), (c^q_2, p^q_2), \ldots, (c^q_n, p^q_n)\}.\end{gathered} \qquad (2)$$

В конечном итоге точность распознавания оценивается по представлению слова первыми альтернативами $\{(c^1_1, p^1_1), (c^1_2, p^1_2), \ldots (c^1_n, p^1_n)\}$. Рассмотрим эвристической алгоритм использования словаря триграмм для пересчета оценок распознавания с целью избавиться от возможных ошибок распознавания, дополнив распознанные символы (2) фиктивными символами c_F.

Для всех символов c^k_i для $1 \le i \le n$, $1 \le k \le q$ (т.е. для всех символов, не являющихся фиктивными) рассмотрим два предшествующих ему символа c^1_{i-2}, c^1_{i-1},

которые считаются уже выбранными, и, используя триграммы $<c^1_{i-2}, c^1_{i-1}, c^k_i>$, содержащие символ c^k_i, вычислим новую оценку $\hat{\theta}(c^k_i)$, которая ранее обозначалась как p^k_i:

$$\hat{\theta}\left(c^k_i\right) = \theta\left(c^k_i\right) \times \theta(c^1_{i-2}, c^1_{i-1}, c^k_i) / \sum_{k=1}^{q} \theta\left(c^k_i\right) \times \theta(c^1_{i-2}, c^1_{i-1}, c^k_i),$$

отсортируем альтернативы по значениям $\hat{\theta}(c^k_i)$ и зафиксируем символ, выбрав первую альтернативу для коррекции следующего символа $c^k_{i+1.}$

Как показывает практика, этот эвристический алгоритм корректировки (ЭАК) дает приемлемые результаты, которые мы будем использовать для сравнения с результатами работы предлагаемого в следующем разделе алгоритма, использующего те же самые исходные данные.

Алгоритм корректировки результатов распознавания с помощью триграмм, основанный на вычислениях маргинальных распределений и байесовских сетей

Рассмотрим другой алгоритм корректировки результатов распознавания с помощью триграмм, основанный на вычислениях маргинальных распределений в вероятностных графических моделях (ВГМ) [5, 6].

Для слова из n символов с каждым знакоместом ассоциируем случайную величину x_i которая принимает значения в конечном пространстве состояний $\mathcal{A}_i$. Без ограничения общности будем считать, что $\forall_i \mathcal{A}_i$ является алфавитом распознавания A. Обозначим через $N = \{1, \ldots, n\}$ множество индексов. Назовем произведение $\mathcal{A} = \times_{i \in N} \cdot \mathcal{A}_i = A^n$ пространством конечных конфигураций вектора $x = (x_i)_{i \in N}$. Далее построим граф $(\mathcal{X}\mathcal{E})$ по следующему правилу [7, 8]. Множество $\mathcal{X} = \{x_i, i \in N\}$ является множеством вершин графа. Будем считать, что с каждой тройкой вершин $<x_j, x_{j+1}, x_{j+2}>$, $i = \overline{1, n-2}$ ассоциирована нормированная функция весов триграмм $\theta(x_j, x_{i+1}, x_{i+2})$. Соединим ребрами вершины из тройки между собой. Совокупность полученных ребер для всех $i = \overline{1, n-2}$ образует множество $\mathcal{E}$. Пример графа для $n = 5$ приведен на *рисунке 1*.

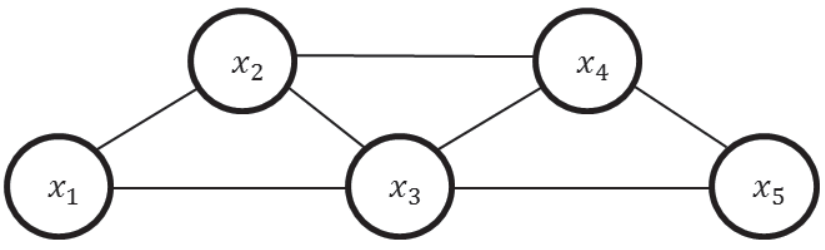


Рис. 1. Пример графа для слова $\{x_1, x_2, x_3, x_4, x_5\}$.

Выпишем формулу для совместного распределения вероятностей

$$p(x_1, \ldots, x_n) = \prod_{i \in N} \theta(x_i) \cdot \prod_{j=1}^{n-2} \theta\left(x_j, x_{j+1}, x_{j+2}\right) / \sum_{\mathcal{A}} \prod_{i \in N} \theta(x_i) \cdot \prod_{j=1}^{n-2} \theta\left(x_j, x_{j+1}, x_{j+2}\right), \quad (3)$$

где $\theta(x_i)$ – функция оценок распознавания для i-го знакоместа (2), $\theta(x_j, x_{j+1}, x_{j+2})$ – нормированная функция весов триграмм.

Используя формулу (3), можно выполнить пересчет оценок. Для этого достаточно посчитать маргинальные распределения:

$$p(x_1) = \sum_{x \setminus \{x_i\}} p(x_1, \ldots, x_n), i \in N. \quad (4)$$

Для вычисления выражения (4) можно применить известный алгоритм HUGIN [8–11]. На первом этапе на основе исходного графа необходимо сформировать дерево сочленений. Такое построение возможно [12] в силу того, что исходный граф является триангулярным. Пусть каждая вершина C_i дерева сочленений представляет собой объединение $\mathcal{X}_i$ таких вершин исходного графа $\mathcal{X}$, что каждые две вершины этого подмножества соединены ребром графа

$$\mathcal{X}^i = \{x_i, x_{i+1}, x_{i+2}\} \subseteq \mathcal{X} i = \overline{1, n-2}.$$

Рассмотрим пересечение двух множеств $\mathcal{X}^i \cap \mathcal{X}^{i+1}$, $i = \overline{1, n-3}$ и назовем данное множество сепаратором S^i. Нетрудно видеть, что $S^i = \{x_{i+1}, x_{i+2}\}$. Соединим сепаратор S^i с вершинами C_i и C_{i+1}. Пример множеств с сепараторами приведен на *рисунке 2*.

Далее, каждой вершине C_i дерева сочленений припишем потенциал, который определяется следующим образом:

$$\phi_{x^i} = \theta\left(x_i, x_{i+1}, x_{i+2}\right) \prod_{x^i \setminus s^{i-1}} \theta_j\left(x_j\right),$$

$i = \overline{1, n-2}$,

с учетом соглашения $S^0 = \{\emptyset\}$.

Кроме этого, положим $\phi_{S^i} = 1, i = \overline{1, n-3}$. Полученная структура является деревом сочленений J = (Т, Ф), где Т – древовидный граф с вершинами, соединенными через сепараторы, и Ф – множество потенциалов, ассоциированных с вершинами дерева и с сепараторами.

Процесс вычисления искомых маргинальных распределений вероятностей будем рассматривать как распространение свидетельства в дереве сочленений [7]. В англоязычной литературе данная процедура называется "belief propagation" [13]. Схематично данный процесс можно представить в виде двух фаз: сбор свидетельств и распространение свидетельств. Сбор свидетельств заключается в передаче сообщений от вершины C_i к вершине C_{i+1}. Математически передача сообщений может быть определена следующим образом:

$$\phi^*_{S^i} = \sum_{x^i \setminus S^i} \phi_{x^i},$$

$$\phi^*_{x^{i+1}} = \phi_{x^{i+1}} \cdot \frac{\phi^*_{S^i}}{\phi_{S^i}}.$$

Распространение свидетельств выполняется в обратную сторону от

Рис. 2. Пример дерева сочленений для слова{x_1, x_2, x_3, x_4, x_5}.

вершины C_{i+1} к вершине C_i в соответствии с аналогичными правилами:

$$\phi^{**}_{S^i} = \sum_{x^{i+1} \setminus S^i} \phi_{x^{i+1}},$$

$$\phi^{**}_{x^i} = \phi^*_{x^i} \cdot \frac{\phi^{**}_{S^i}}{\phi^*_{S^i}}.$$

Следует заметить, что $\phi^{**}_{x^{n-2}} = \phi^*_{x^{n-2}}$.

В результате работы алгоритма получаются значения для скорректированных оценок, представляющих собой векторную функцию. Выражение для скорректированной оценки можно записать следующим образом:

$$p(x_j) \propto \sum_{x^i \setminus \{x_j\}} \phi^{**}_{x^i}, i = \overline{1, n-2}, j = \overline{i, i+2}.$$

Экспериментальные данные

В экспериментах для построения словарей триграмм использовали три набора данных: набор фамилий, имен и отчеств граждан Российской Федера-

Таблица 1. *Результаты работы алгоритмов на тестовом наборе T_1*

Оценка результатов распознавания	Название текстового поля (количество образцов)					
	Фамилия (896)		Имя (915)		Отчество (911)	
	Алгоритм					
	ЭАК	ВГМ	ЭАК	ВГМ	ЭАК	ВГМ
Ошибки в исходном распознавании, не исправленные триграммами	44 (4.91%)	38 (4.24%)	28 (3.06%)	30 (3.28%)	31 (3.40%)	21 (2.31%)
Ошибки в исходном распознавании, исправленные триграммами	36 (4.02%)	42 (4.69%)	38 (4.15%)	36 (3.93%)	48 (5.27%)	58 (6.37%)
Ошибки триграмм, привнесенные в правильные результаты	5 (0.56%)	1 (0.11%)	9 (0.98%)	4 (0.44%)	5 (0.55%)	1 (0.11%)
Ошибки в исходном распознавании и при использовании триграмм отсутствуют	811 (90.51%)	815 (90.96%)	840 (91.80%)	845 (92.35%)	827 (90.78%)	831 (91.22%)
Точность исходного распознавания	816 (91.07%)	816 (91.07%)	849 (92.79%)	849 (92.79%)	832 (91.33%)	832 (91.33%)
Точность с триграммами	847 (94.53%)	857 (95.65%)	878 (95.96%)	881 (96.28%)	875 (96.05%)	889 (97.59%)
Выигрыш в точности при использовании триграмм	3.46%	4.58%	3.17%	3.50%	4.72%	6.26%

Таблица 2. *Результаты работы алгоритмов на тестовом наборе* T_2

Оценка результатов распознавания	Название текстового поля (количество образцов)					
	Фамилия (776)		Имя (794)		Отчество (784)	
	Алгоритм					
	ЭАК	ВГМ	ЭАК	ВГМ	ЭАК	ВГМ
Ошибки в исходном распознавании, не исправленные триграммами	16 (2.06%)	12 (1.55%)	7 (0.88%)	5 (0.63%)	9 (1.15%)	8 (1.02%)
Ошибки в исходном распознавании, исправленные триграммами	21 (2.71%)	25 (3.22%)	24 (3.02%)	26 (3.27%)	26 (3.32%)	27 (3.44%)
Ошибки триграмм, привнесенные в правильные результаты	3 (0.39%)	2 (0.26%)	3 (0.38%)	1 (0.13%)	2 (0.26%)	0 (0.0%)
Ошибки в исходном распознавании и при использовании триграмм отсутствуют	736 (94.85%)	737 (94.97%)	760 (95.72%)	762 (95.97%)	747 (95.28%)	749 (95.54%)
Точность исходного распознавания	739 (95.23%)	739 (95.23%)	763 (96.10%)	763 (96.10%)	749 (95.54%)	749 (95.54%)
Точность с триграммами	757 (97.55%)	762 (98.20%)	784 (98.74%)	788 (99.24%)	773 (98.60%)	776 (98.98%)
Выигрыш в точности при использовании триграмм	2.32%	2.96%	2.64%	3.15%	3.06%	3.44%

ции, объем каждого из которых составлял примерно 1 500 000 примеров слов.

Построенные словари испытывали на двух тестовых наборах: T_1 (низкое качество оцифровки) и T_2 (среднее качество оцифровки). Каждый пример из тестовой выборки представляет собой строку, содержащую известные символы поля и набор пар (c_1, p_1), (c_2, p_2), …, (c_n, p_n) с результатами распознавания.

Для алгоритма ЭАК и алгоритма ВГМ, описанных выше, были получены результаты, приведенные в *таблицах 1* и *2*.

Из таблиц следует, что на наборах T_1 и T_2 оба алгоритма существенно улучшают исходные результаты распознавания. При этом алгоритм ВГМ дает лучшие результаты, чем алгоритм ЭАК.

Экспериментальные данные показывают, что в некоторых случаях использование триграмм приводит к ухудшению верных результатов распознавания. Как правило, такие случаи объясняются двумя причинами:

- исходное распознавание дает не очень уверенный результат с низкими оценками первой альтернативы, а триграммы, получаемые из прочих альтернатив, имеют больший вес (например в слове «ДИНАР» знакоместо с символом «Д» имеет низкую оценку, равную 0.57, что приводит к замене на символ «Л» из-за приблизительного равенства веса триграммы «ДИН» – 0.0001, тогда как вес триграммы «ЛИН» примерно равен 0.0034);
- в словаре триграмм отсутствуют триграммы, необходимые для подтверждения троек символов в редких именах, фамилиях или отчествах (например в словаре триграмм не содержится триграммы «ЗРВ», отчего слово «АЗРВАРД» заменяется на «АЗРААРД»).

Выводы

В данной работе, посвященнной исследованию влияния моделей языка на основе символов триграмм на точность распознавания, были предложены два алгоритма: модель на основе цепей Маркова (алгоритм ЭАК) и алгоритм на основе вероятностной графической модели (алгоритм ВГМ). Вычисления маргинальных распределений в вероятностной графической модели выполнялись с применением алгоритма HUGIN. Для оценки предложенных подходов использовались два тестовых набора документов с разным качеством оцифровки. Проведенные эксперименты показывают улучшение качества распознавания в условиях применения моделей символов триграмм. При низком качестве оцифровки точность распознавания может быть повышена более чем на 6%.

В целом алгоритм ВГМ повышает точность распознавания лучше, чем алгоритм ЭАК, но при этом сложность алгоритма ВГМ выше, чем сложность алгоритма ЭАК.

Литература

1. ***Л.П. Гниловская, Н.Ф. Гниловская***
Культура народов Причерноморья, 2004, №48, Т. 2, 171.
2. ***K. Kukich***
ACM Computing Surveys (CSUR), 1992, **24**(4), 377.
DOI: 10.1145/146370.146380.
3. ***В.М. Кляцкин, Н.В. Котович, О.А. Славин***
Труды ИСА РАН, 2009, **45**, 260.
4. ***О.А. Славин, И.М. Яншевский***
Труды ИСА РАН, 2012, **62**(2), 30.
5. ***R. Klinger, K. Tomanek***
Classical Probabilistic Models and Conditional Random Fields: Algorithm Engineering Report, TR07-2-013, Dep. Computer Science, Dortmund University of Technology, 2007, 31 pp. (http://ls11-www.cs.uni-dortmund.de/_media/techreports/tr07-13.pdf).
6. ***R. Dechter***
Synthesis Lectures on Artificial Intelligence and Machine Learning, 2013, **7**(3), 1. DOI: 10.2200/S00529ED1V01Y201308AIM023.
7. ***L.E. Sucar***
Probabilistic Graphical Models: Principles and Applications, Ser. Advances in Computer Vision and Pattern Recognition, UK, London, Publ. Springer-Verlag London, 2015, 253 pp.
DOI: 10.1007/978-1-4471-6699-3.
8. ***R. Cowell, P. Dawid, S. Lauritzen, D. Spiegelhalter***
Probabilistic Networks and Expert Systems, USA, NY, New York, Springer-Verlag New York Inc., 1999, 322 pp. DOI: 10.1007/b97670.
9. ***S.L. Lauritzen, D.J. Spiegelhalter***
J. R. Statist. Soc. B (Methodological), 1988, **50**(2), 157.
10. ***V. Lepar, P. Shenoy***
B *Proc. 14th Conf. Annual Conference on Uncertainty in Artificial Intelligence (UAI-98) (USA, Wisconsin, Madison, 24–26 July, 1998),* USA, CA, San Francisco, Morgan Kaufmann Publ., 1998, pp. 328–337.
11. ***S.L. Lauritzen, F.V. Jensen***
Annals of Mathematics and Artificial Intelligence, 1997, **21**(1), 51.
DOI: 10.1023/A:1018953016172.
12. ***T. Schmidt, P.P. Shenoy***
Artificial Intelligence, 1998, **102**(2), 323.
DOI: 10.1016/S0004-3702(98)00047–2.
13. ***Michael I. Jordan***
An Introduction to Probabilistic Graphical Models, Manuscript used for Class Notes of CS281A at UC Berkeley, Fall 2002, 2002, 102 pp. (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.116.7467&rep=rep1&type=pdf).
14. ***P.P. Shenoy, G. Shafer***
B *Classic Works of the Dempster-Shafer Theory of Belief Functions. Ser. Studies in Fuzziness and Soft Computing,* Eds R.R. Yager, L. Liu, 1990, **219**, 499. DOI: 10.1007/978-3-540-44792-4_20.

English

N-Grams Algorithm Application for the Correction of Recognition Results*

Temujin V. Manzhikov –
National University of Science and Technology MISIS
4, Leninskiy Ave.,
Moscow, 119049, Russia
e-mail: tmanzhikov@gmail.com

Oleg A. Slavin –
Institute for System Analysis FRC "Computer Science and Control" RAS
9, 60-letiya Octyabrya Ave.,
Moscow, 117312, Russia
e-mail: oslavin@isa.ru

Igor A. Faradzhev –
Institute for System Analysis FRC "Computer Science and Control" RAS
9, 60-letiya Octyabrya Ave.,
Moscow, 117312, Russia
e-mail: ifardjev@yahoo.com

Igor M. Janiszewski –
Institute for System Analysis FRC "Computer Science and Control" RAS
9, 60-letiya Octyabrya Ave.,
Moscow, 117312, Russia
e-mail: igor_y@cs.isa.ru

Abstract

The authors investigated the N-grams use for the correction of the recognition results of images documents through the example of text fields of a passport of the Russian Federation citizen. For the trigrams, the two algorithms for adjusting the recognition results are suggested. One of the algorithms is based on the use of trigram probabilities combined with recognition estimates, which are also interpreted as probabilities. The second algorithm is built on the definition of marginal distributions and the calculations on graphs based on the Bayesian networks. The results of the experiments on the algorithms usage and comparison of the both algorithms characteristics are presented in the article.

Keywords: character recognition, N-grams, trigram, marginal distribution, calculation on graphs.

* *The work was financially supported by RFBR (projects 13-07-12170, 13-07-12171, 16-07-01051).*

Images & Tables

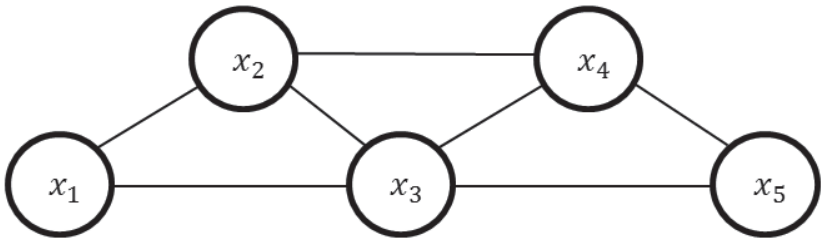


Fig. 1. Sample of graph for word $\{x_1, x_2, x_3, x_4, x_5\}$.

Fig. 2. Sample of graph tree for word $\{x_1, x_2, x_3, x_4, x_5\}$.

Table 1. *The algorithms performance at work with the dataset* T_1

Evaluation of the recognition results	Text field name (number of samples)					
	Surname (896)		Name (915)		Patronym (911)	
	Algorithm					
	EAK	VGM	EAK	VGM	EAK	VGM
OCR errors not corrected with trigrams use	16 (2.06%)	12 (1.55%)	7 (0.88%)	5 (0.63%)	9 (1.15%)	8 (1.02%)
OCR errors corrected using trigrams	21 (2.71%)	25 (3.22%)	24 (3.02%)	26 (3.27%)	26 (3.32%)	27 (3.44%)
Errors introduced by trigrams into originally correct OCR results	3 (0.39%)	2 (0.26%)	3 (0.38%)	1 (0.13%)	2 (0.26%)	0 (0.0%)
Originally correct OCR results, not changed with trigrams	736 (94.85%)	737 (94.97%)	760 (95.72%)	762 (95.97%)	747 (95.28%)	749 (95.54%)
Original OCR precision	739 (95.23%)	739 (95.23%)	763 (96.10%)	763 (96.10%)	749 (95.54%)	749 (95.54%)
OCR precision after trigrams use	757 (97.55%)	762 (98.20%)	784 (98.74%)	788 (99.24%)	773 (98.60%)	776 (98.98%)
OCR precision improvement when using trigrams	2.32%	2.96%	2.64%	3.15%	3.06%	3.44%

DOI: 10.22204/2410-4639-2016-092-04-116-123 № 4 (92) октябрь–декабрь 2016 г.

Table 2. *The algorithms performance at work with the dataset* T_2

Evaluation of the recognition results	Text field name (number of samples)					
	Surname (776)		Name (794)		Patronym (784)	
	Algorithm					
	EAK	VGM	EAK	VGM	EAK	VGM
OCR errors not corrected with trigrams use	16 (2.06%)	12 (1.55%)	7 (0.88%)	5 (0.63%)	9 (1.15%)	8 (1.02%)
OCR errors corrected using trigrams	21 (2.71%)	25 (3.22%)	24 (3.02%)	26 (3.27%)	26 (3.32%)	27 (3.44%)
Errors introduced by trigrams into originally correct OCR results	3 (0.39%)	2 (0.26%)	3 (0.38%)	1 (0.13%)	2 (0.26%)	0 (0.0%)
Originally correct OCR results, not changed with trigrams	736 (94.85%)	737 (94.97%)	760 (95.72%)	762 (95.97%)	747 (95.28%)	749 (95.54%)
Original OCR precision	739 (95.23%)	739 (95.23%)	763 (96.10%)	763 (96.10%)	749 (95.54%)	749 (95.54%)
OCR precision after trigrams use	757 (97.55%)	762 (98.20%)	784 (98.74%)	788 (99.24%)	773 (98.60%)	776 (98.98%)
OCR precision improvement when using trigrams	2.32%	2.96%	2.64%	3.15%	3.06%	3.44%

References

1. ***L.P. Gnilovskaya, N.F. Gnilovskaya***
   Culture of Peoples of the Black Sea Region [Kultura narodov Prichernomorya], 2004, №48, Vol. 2, 171 (in Russian).
2. ***K. Kukich***
   ACM Computing Surveys (CSUR), 1992, **24**(4), 377. DOI: 10.1145/146370.146380.
3. ***V.M. Klyatskin, N.V. Kotovich, O.A. Slavin***
   Proc. ISA RAS [Trudy Instituta sistemnogo analiza RAN], 2009, **45**, 260 (in Russian).
4. ***O.A. Slavin, I.M. Janiszewski***
   Proc. ISA RAS [Trudy Instituta sistemnogo analiza RAN], 2012, **62**(2), 30 (in Russian).
5. ***R. Klinger, K. Tomanek***
   Classical Probabilistic Models and Conditional Random Fields: Algorithm Engineering Report, TR07-2-013, Dep. Computer Science, Dortmund University of Technology, 2007, 31 pp. (http://ls11-www.cs.uni-dortmund.de/_media/techreports/tr07-13.pdf).
6. ***R. Dechter***
   Synthesis Lectures on Artificial Intelligence and Machine Learning, 2013, **7**(3), 1. DOI: 10.2200/S00529ED1V01Y201308AIM023.
7. ***L.E. Sucar***
   Probabilistic Graphical Models: Principles and Applications, Ser. Advances in Computer Vision and Pattern Recognition, UK, London, Publ. Springer-Verlag London, 2015, 253 pp. DOI: 10.1007/978-1-4471-6699-3.
8. ***R. Cowell, P. Dawid, S. Lauritzen, D. Spiegelhalter***
   Probabilistic Networks and Expert Systems, USA, NY, New York, Springer-Verlag New York Inc., 1999, 322 pp. DOI: 10.1007/b97670.
9. ***S.L. Lauritzen, D.J. Spiegelhalter***
   J. R. Statist. Soc. B (Methodological), 1988, **50**(2), 157.
10. ***V. Lepar, P. Shenoy***
    In *Proc. 14th Conf. Annual Conference on Uncertainty in Artificial Intelligence (UAI-98) (USA, Wisconsin, Madison, 24–26 July, 1998),* USA, CA, San Francisco, Morgan Kaufmann Publ., 1998, pp. 328–337.
11. ***S.L. Lauritzen, F.V. Jensen***
    Annals of Mathematics and Artificial Intelligence, 1997, **21**(1), 51. DOI: 10.1023/A:1018953016172.
12. ***T. Schmidt, P.P. Shenoy***
    Artificial Intelligence, 1998, **102**(2), 323. DOI: 10.1016/S0004-3702(98)00047–2.
13. ***Michael I. Jordan***
    An Introduction to Probabilistic Graphical Models, Manuscript used for Class Notes of CS281A at UC Berkeley, Fall 2002, 2002, 102 pp. (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.116.7467&rep=rep1&type=pdf).
14. ***P.P. Shenoy, G. Shafer***
    In *Classic Works of the Dempster-Shafer Theory of Belief Functions. Ser. Studies in Fuzziness and Soft Computing,* Eds R.R. Yager, L. Liu, 1990, **219**, 499. DOI: 10.1007/978-3-540-44792-4_20.

Многовариантное численное моделирование при решении задачи исследования устойчивости методов статистического распознавания к искажениям образов *

Б.М. Гавриков, Н.В. Пестрякова

Описывается вероятностный способ классификации, в основе которого лежит полиномиальная регрессия. Метод используется для распознавания печатных и рукопечатных символов. Посредством многовариантного численного моделирования показано, что метод устойчив к искажениям исходных образов.

Ключевые слова: классификация, полиномиальная регрессия, печатные и рукопечатные символы.

** Работа выполнена при финансовой поддержке РФФИ (проекты №№ 13-07-00262 и 13-07-12176).*

Введение

К настоящему времени разработано большое количество способов классификации объектов разнообразной этимологии. Зачастую изучение свойств этих методов сводилось лишь к сопоставлению с другими подходами по таким характеристикам, как точность и быстродействие [1]. В последние годы были предложены методики анализа монотонности (надежности) классификаторов [2]. Большое значение имеет устойчивость метода к искажениям исходных образов (как символьных, так и отличных от них) [3].

Однако вне области исследований оказалась внутренняя структура самого «устройства» для распознавания. Оно фактически используется в качестве «черного ящика», работающего следующим образом: на его вход поступает кодированное изображение объекта, а на выходе получается номер класса, к которому этот объект относится, а также, возможно, оценка распознавания. О способе получения этой оценки имеется самое общее представление.

Для описываемого в настоящей статье метода распознавания таким устройством с неизученной структурой является двумерная матрица весовых коэффициентов. Наше знание об оценке распознавания ограничивалось пониманием алгоритма построения механизма для ее вычисления, а также непосредственной реализацией этого механизма на компьютере.

Требовалось провести исследование структуры матрицы весовых коэффициентов, используемой при распознавании, а также статистических свойств функции оценки на всем обучающем множестве.

Рассматриваемая проблематика является весьма актуальной, поскольку результаты подобного анализа можно использовать, в частности, для оптимизации матрицы распознавания. Заметим, что пока еще отсутствует некий стандартный способ проведения такого рода исследований. В настоящей работе описаны численные подходы к решению указанной задачи. При этом используется метод распознавания символов, разработанный авторами [4]. Он имеет ряд достоинств, а именно: точность, быстродействие, надежность генерируемых им оценок (монотонность), устойчивость к искажениям. Метод является вероятностным, поскольку в его основе лежит восстановленный с большой степенью достоверности некоторый неизвестный вероятностный закон, в соответствии с которым распределены элементы обучающей последовательности символов, моделирующей датчик случайных векторов. Высокий уровень точности распознавания на обучающем множестве служит подтверждением достоверности этого приближения.

С применением численного подхода проведено исследование внутреннего устройства механизма распознавания. При этом для обучения и

ГАВРИКОВ
Борис Михайлович
Институт системного анализа
ФИЦ «Информатика и управление» РАН

ПЕСТРЯКОВА
Надежда Владимировна
Институт системного анализа
ФИЦ «Информатика и управление» РАН

распознавания использовалась база рукопечатных цифр, представляемых заданным способом в виде векторов признакового пространства. Это множество, имеющее большой объем, набор его элементов является случайным. Распознавались также искаженные изображения обучающих символов, полученные на основе разработанной авторами математической модели.

Метод распознавания

Алгоритм позволяет по предъявляемому растру изображения определить, какому символу из некоторого конечного множества с K элементами он соответствует. Представлением символа является растр, состоящий из $N = N1 \times N2$ серых пикселов. Пронумеровав все пикселы растра, запоминаем в i-й компоненте ($1 \le i \le N$) вектора $v \in R^N$ состояние i-го пиксела (яркость), а именно значение на отрезке [0,1] для серого растра. Пусть $V = \{v\}$ – совокупность всевозможных растров. Очевидно, $V \subseteq R^N$, причем поскольку пикселы серые, то $V = [0,1]^N$ – N-мерный единичный куб в R^N. R^N – N-мерное арифметическое векторное пространство.

Отождествим k-й символ с базисным вектором $e_k = (0\ldots1\ldots0)$ (1 на k-м месте, $1 \le k \le K$) из R^K. Обозначим $Y = \{e_1,\ldots,e_K\}$.

Пусть для предъявляемого растра $v \in V$ можно найти $p_k(v)$ – вероятность того, что растр изображает символ с номером k, $1 \le k \le K$. Тогда результатом распознавания считается символ с порядковым номером k_o, где

$$p_{ko}(v) = \max p_k(v),\ 1 \le k \le K. \quad (1)$$

Для решения задачи требуется вычислить вектор вероятностей $(p_1(v), p_2(v),\ldots, p_K(v))$. Он может быть найден на основе метода наименьших квадратов [5]. Приближенные значения компонент $(p_1(v),\ldots,p_K(v))$ в соответствии с методом полиномиальной регрессии следует искать в виде разложения по произведениям степеней яркости в различных пикселах:

$$p_k(v) \cong c_0^{(k)} + \sum_{i=1}^{N} c_i^{(k)} v_i + \sum_{i,j=1}^{N} c_{i,j}^{(k)} v_i v_j + \ldots, \quad (2)$$

где $1 \le k \le K$, v_i – значение яркости в пикселе с номером i.

Суммы в уравнении (2) конечные и определяются выбором базисных мономов. Вводится в рассмотрение полиномиальный вектор $x(v) = (1, v_1, \ldots, v_N, \ldots)^T$ конечной размерности L из приведенных в формуле (2) базисных мономов, упорядоченных некоторым образом. Здесь и далее T означает транспонированный вектор или матрица. Тогда уравнение (2) можно записать:

$$p(v) = (p_1(v),\ldots,p_K(v))^T \cong A^T x(v). \quad (3)$$

Столбцами матрицы A размера $L \times K$ являются векторы $a^{(1)},\ldots, a^{(K)}$. Каждый такой вектор составлен из коэффициентов при мономах соответствующей строки уравнение (2) с совпадающим индексом k, упорядоченных так, как это сделано в $x(v)$.

Для вычисления матрицы A используется обучающая выборка: $[v^{(1)},y^{(1)}], [v^{(2)},y^{(2)}],\ldots, [v^{(J)},y^{(J)}]$.

Вектор $v^{(j)}$ соответствует растру изображения. Вектор $y^{(j)}$ кодирует символ. Все его компоненты нулевые, кроме той, номер которой соответствует номеру символа, – она равна 1.

$y^{(j)} = (0, \ldots, 1, \ldots, 0)$.

Приближенное значение A находят следующим образом (обучение):

$$A \cong \left(\frac{1}{J}\sum_{j=1}^{J} x^{(j)}(x^{(j)})^T\right)^{-1}\left(\frac{1}{J}\sum_{j=1}^{J} x^{(j)}(y^{(j)})^T\right). \quad (4)$$

Проблема обращения заполненной матрицы большой размерности до сих пор не решена [6]. Чтобы обойти ее, предлагается [7] правую часть уравнения (4) вычислить с использованием рекуррентной процедуры:

$$A_j = A_{j-1} - \alpha_j G_j x^{(j)}[A_{j-1}^T x^{(j)} - y^{(j)}]^T, \quad (5)$$

где $\alpha_j = 1/J$.

$$G_j = \frac{1}{1-\alpha_j}\left[G_{j-1} - a_j \frac{G_{j-1} x^{(j)}(x^{(j)})^T G_{j-1}}{1 + \alpha_j((x^{(j)})^T G_{j-1} x^{(j)} - 1)}\right],$$

где $1 \le j \le J$, A_0 и G_0 заданы некоторым образом, G_j – матрица размера $L \times L$.

Используется упрощенная модификация процедуры (5):

$$G_j \equiv D^{-1},\ D = \mathrm{diag}(E\{x_1^2\}, E\{x_2^2\},\ldots, E\{x_L^2\}), \quad (6)$$

где $x_1, x_2,\ldots,x_L$ – компоненты вектора $x(v)$, E – математическое ожидание. В этом случае были получены приемлемые практические результаты для серых

растров размера $N = 256 = 16\times16$.

Для рукопечатных цифр вектор x имеет вид:

$$\mathrm{x} = (1, \{v_i\}, \{v_i^2\}, \{(\delta v_i)_r\}, \{(\delta v_i)_r^2\}, \{(\delta v_i)_y\}, \{(\delta v_i)_y^2\}, \{(\delta v_i)_r^4\}, \{(\delta v_i)_y^4\}, \{(\delta v_i)_r(\delta v_i)_y\}, \{(\delta v_i)_r^2(\delta v_i)_y^2\}, \{(\delta v_i)_r^4(\delta v_i)_y^4\}, \{(\delta v_i)_r((\delta v_i)_r)_l\}, \{(\delta v_i)_y((\delta v_i)\mathrm{y})_l\}, \{(\delta v_i)_r((\delta v_i)_y)_l\}, \{(\delta v_i)_y((\delta v_i)_r)_l\}, \{(\delta v_i)_r((\delta v_i)_r)_d\}, \{(\delta v_i)_y((\delta v_i)_y)_d\}, \{(\delta v_i)_r((\delta v_i)_y)_d\}, \{(\delta v_i)_y((\delta v_i)_r)_d\}). \quad (7)$$

Для печатных цифр вектор x составлен из элементов в первой строке (7):

$$\mathrm{x} = (1, \{v_i\}, \{v_i^2\}, \{(\delta v_i)_r\}, \{(\delta v_i)_r^2\}, \{(\delta v_i)_y\}, \{(\delta v_i)_y^2\}). \quad (8)$$

Через $(\delta v_i)_r$ и $(\delta v_i)_y$ обозначены конечные центральные разности величин v_i по ортогональным направлениям ориентации растра – нижние индексы r и y соответственно.

Компоненты вектора x, не имеющие индекса l (left) или d (down), вычисляются для всех пикселов растра, с индексом l – кроме левых граничных, с индексом d – кроме нижних граничных. Индекс l или d означает, что величины относятся к пикселу слева или снизу от данного. Вне растра $v_i = 0$ (используется при вычислении конечных разностей на границе растра).

Поскольку рукопечатные символы имеют меньшую толщину, чем печатные, использовался прием искусственного «уширения» изображения. Суть его заключается в том, что при вычислении компонент полиномиального вектора имеющееся изображение искусственно увеличивается в размере на один пиксел в направлении, ортогональном соответствующему участку границы.

При вычислении элементов матрицы D (7) для каждого j-го элемента базы символов строится вектор x^j согласно (7) или (8). Попутно рассчитываются компоненты вспомогательного вектора m^j по формуле:

$$m_p^j = (1\text{-}1/j)\, m_p^{j-1} + (\mathrm{x}_p^j)^2 /j, \quad (9)$$

где $j = 1,\ldots,J$, $\quad p = 1,\ldots,L$.

В конце этой процедуры для последнего элемента имеем согласно (6):

$$G_j \equiv D^{-1} = \mathrm{diag}\,(1/m_1^J, 1/m_2^J, \ldots, 1/m_L^J). \quad (10)$$

После вычисления G_j для каждого j-го элемента базы символов строится вектор x^j согласно (7) или (8) и находятся элементы матрицы A_j (5):

$$a_j^{pk} = a_{j-1}^{pk} - \alpha_j x_p^j \Big(\sum_{i=1}^{L} a_{j-1}^{ik} x_i^j - y_k^i\Big) / m_p^J, \quad (11)$$

где $\alpha_j = 1/J$,

$A_j = [a_j^{pk}]$,

где $j = 1,\ldots,J$, $p = 1,\ldots,L$, $k = 1,\ldots,K$.

При распознавании по изображению строится вектор x согласно (7) или (8). Далее по формуле (3), используя $A = A_j$ (11), вычисляют оценки для каждого из символов. Затем находят символ с максимальной оценкой.

Получаемые из-за приближенности метода отрицательные оценки искусственно обнуляли, а превышающие единицу делали равными 1.

Моделирование искажений изображений символов

Для исследования устойчивости метода распознавания к искажениям изображений в качестве базы распознавания использована заданная модификация базы обучения. Рассмотрены две модели искажения, характеризуемые увеличением яркости в пикселах (затемнение), а также ее уменьшением (засветление).

Моделирование процесса нарастания различия обучающего и распознаваемого множеств проводится следующим образом. При затемнении на этапе распознавания значение яркости для каждого пиксела растра постепенно увеличивается: $v_i \rightarrow v_i + 0.01 \cdot n$, где $n = 0, 1, .., 100$. Если для каких-то пикселов начиная с некоторого n имеем $v_i > 1$, то считаем $v_i = 1$. При засветлении яркость в пикселах уменьшается с ростом n: $v_i \rightarrow v_i - 0.01 \cdot n$, где $n = 0, 1, \ldots, 100$. Если получено $v_i < 0$, то считаем $v_i = 0$. Будем называть n степенью затемнения или засветления.

На *рисунке 1* приведена зависимость от n доли (в процентном выражении) числа нераспознанных изображений относительно их общего числа (*mis%t* – затемнение, *mis%s* – засветление). Очевидно, что увеличение n, соответствующее нарастанию искажения изображений символов, должно приводить к уменьшению числа правильных распознаваний. Однако следует от-

DOI: 10.22204/2410-4639-2016-092-04-124-134 № 4 (92) октябрь–декабрь 2016 г.

метить, что особенности динамики этого процесса существенно зависят как от типа написания, так и от направления изменения яркости (ее усиления или ослабления). Для рукопечатных цифр указанная величина при затемнении растет монотонно, причем на отрезке $24 \le n \le 32$ темпы ее роста стремительно увеличиваются и в дальнейшем остаются высокими. При засветлении рукопечатных символов в целом наблюдается картина монотонного роста, но на участке $64 \le n \le 72$ имеется небольшая немонотонность. Для печатных *mis%t* и *mis%s* увеличиваются монотонно, причем процесс затемнения характеризуется резким увеличением темпов роста величины *mis%t* при $48 \le n \le 56$, которые при больших значениях n еще увеличиваются, аналогично наблюдавшемуся для рукопечатных. Для любого типа написания при засветлении соответствующая кривая является более пологой, чем при затемнении. Засветление печатных образов приводит к наименьшему нарастанию доли неправильно распознанных образов по сравнению с засветлением рукопечатных, а также затемнением и печатных, и рукопечатных символов: при $n = 96$ для печатных цифр *mis%s* = 2.4%, *mis%t* = 82.9%, для рукопечатных *mis%t* = 81.0%, *mis%s* = 46.1%.

Для проведения последующего анализа следует заметить, что важнейшим показателем «правильного» поведения метода является уменьшение оценки при нарастании различия обучающего и распознаваемого множеств.

Средняя оценка распознавания для рукопечатных символов и при затемнении (*Prb_t*) и при засветлении (*Prb_s*) сначала уменьшается, а затем увеличивается *(рис. 2)*. Точка минимума определяет предел достоверности оценок. Для *Prb_t* средние темпы и падения, и роста в 1.5-2 раза выше, чем для *Prb_s*. Необходимо отметить, что минимум *Prb_t* достигается при $n = 32$, а минимум *Prb_s* – при $n = 64$, то есть именно на участках особенного поведения величин *mis%t* и *mis%s*. Для печатных цифр *Prb_s* стремительно монотонно падает, а *Prb_t* при общей тенденции к гораздо более медленному монотонному уменьшению имеет небольшой участок немонотонности (ограничение предела достоверности оценок значением $n = 72$ определяет локальный максимум при $n = 80$); ранее отмечалось, что здесь также нарушается гладкость для *mis%t*. Итак, пределу достоверности оценки для рукопечатных символов соответствует значение $n = 32$ при затемнении и $n = 64$

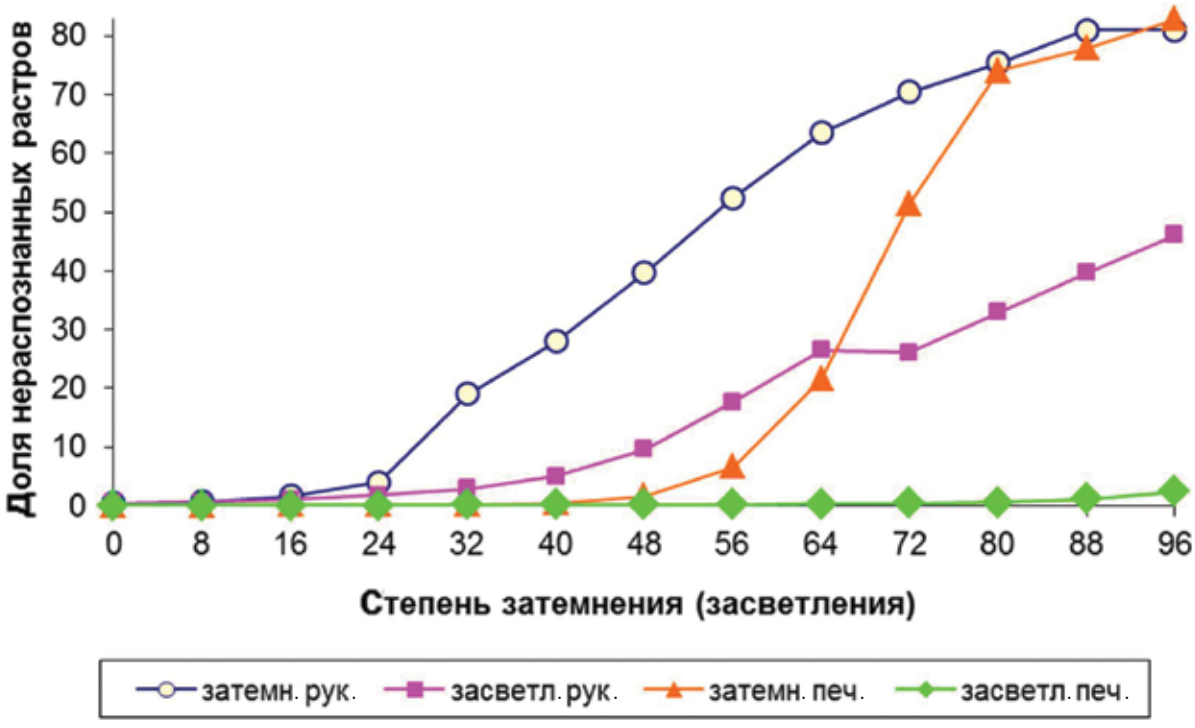


Рис. 1. *Доля нераспознанных растров при затемнении и засветлении.*

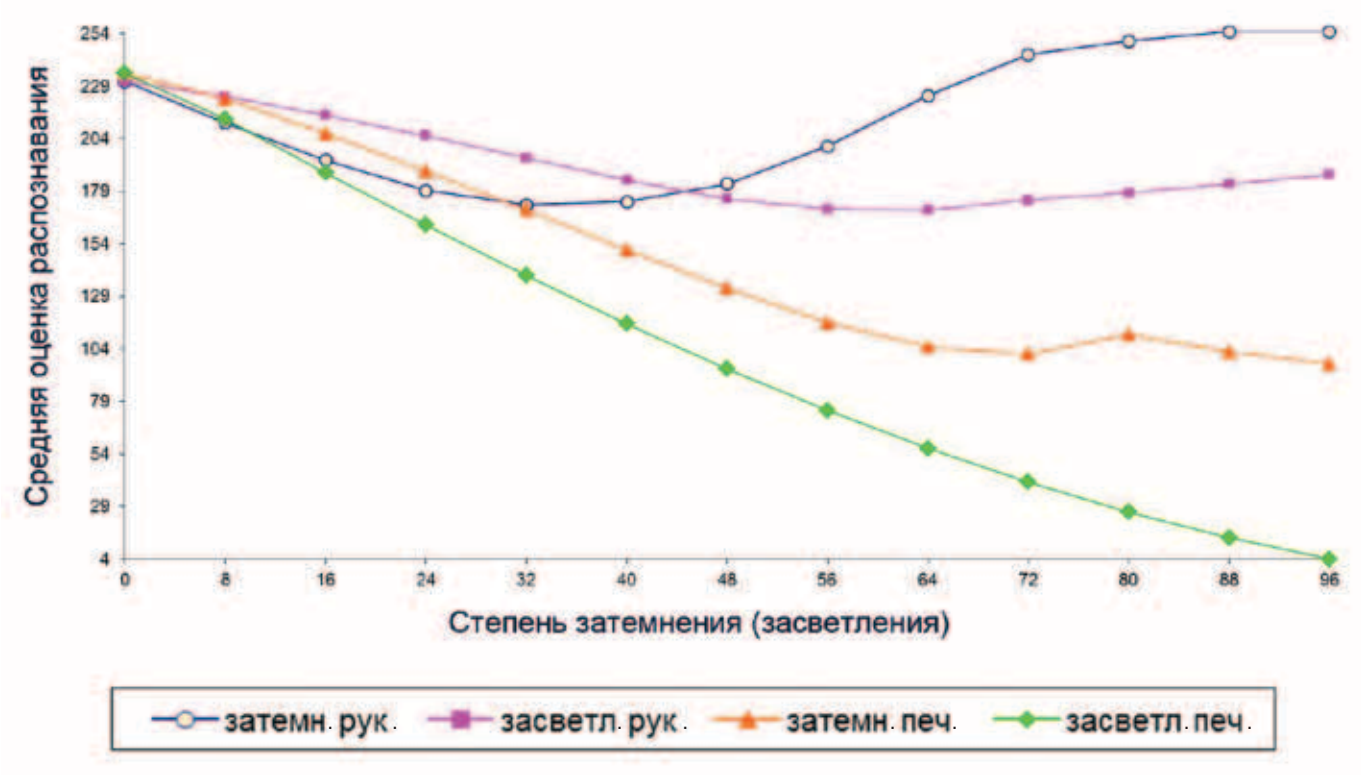


Рис. 2. Средняя оценка распознавания при затемнении и засветлении.

при засветлении, в то время как для печатных при затемнении ему соответствует значение $n = 72$, а при засветлении оценка является «правильной» вплоть до максимальных значений.

Средняя оценка при $n = 0$ для печатных цифр несколько больше, чем для рукопечатных. Но при засветлении печатных изображений *Prb_s* резко падает. *Prb_s* при $n > 0$ меньше *Prb_t* для печатных, а при $n = 3$ становится ниже оценки засветления и при $n = 12$ – затемнения рукопечатных. Тем не менее, несмотря на падение оценки распознавания при засветлении печатных символов, точность уменьшается медленно. Все это свидетельствует о высокой устойчивости метода к распознаванию печатных символов при уменьшении яркости, а также о достоверности выставляемых оценок.

Анализ структуры матрицы весовых коэффициентов

Как обучение, так и распознавание проводилось на базе рукопечатных цифр из 174778 элементов. Пусть на вход распознавателя поступают перенумерованные изображения ($1 \le j \le J_{k0}$) только одного определенного k_0-го символа (цифр «0», «1», …, «9»). Каждое из них будет распознано в качестве k-го символа с оценкой p^j_k для $1 \le k \le K$, где $K = 10$. Согласно (2) и (3)

$$p^j_k = \sum_{l=1}^{L} a^{lk} x^j_l , \qquad (12)$$

где $1 \le j \le J_{k0}$, $k = 1,\ldots,K$, a^{lk} - весовой коэффициент, x^j_i – моном.

Определим, в каком диапазоне лежат значения весовых коэффициентов a^{lk}, входящие в виде множителей в слагаемые $a^{lk}\ x^j_l$ при вычислении оценки распознавания, а также как они распределены внутри этого диапазона. Для вычисления оценок различных символов как альтернатив распознавания, рассматриваемых по отдельности (фиксированные $k = 1,\ldots,K$), диапазон совокупности указанных весовых коэффициентов несколько отличается. Диапазон, в котором лежат значения всех весовых коэффициентов a^{lk} соответствующей матрицы, определяется двумя значениями:

$$\min_{l,k}(a^{lk}) = -0.376, \max_{l,k}(a^{lk}) = 0.311. \qquad (13)$$

Делим этот диапазон $[\min_{l,k}(a^{lk}), \max_{l,k}(a^{lk})]$ на 10 равных частей и для

 DOI: 10.22204/2410-4639-2016-092-04-124-134 № 4 (92) октябрь–декабрь 2016 г.

различных входящих символов по отдельности определяем, какое количество указанных весовых коэффициентов оказывается на каждом из этих отрезков.

Распределение числа рассматриваемых слагаемых, включая расположение отрезков с их максимальным количеством, оказывается полностью синхронным для различных символов, причем максимум лежит в окрестности нуля. По обе стороны этого отрезка имеется небольшое количество (0.9% – 3.4%) таких слагаемых, причем их число уменьшается к концам этого диапазона. Для каждого символа суммарное количество весовых коэффициентов равно L.

Статистический анализ функции оценки на обучающем множестве

Для изображений любого отдельно взятого распознаваемого символа определим, как значения $a^{lk0}\,x_l^j$ количественно распределены внутри объединенного для всех символов диапазона изменения этой величины в случае, когда рассматривается единственная альтернатива распознавания, а именно только сам этот символ ($k = k_0$). При такой постановке задачи нижняя и верхняя границы соответствующего диапазона вычисляются следующим образом:

$$f = \min_{k0} \min_{l,j}(a^{lk0}x_l^j),$$
$$F = \min_{k0} \min_{l,j}(a^{lk0}x). \qquad (14)$$

Этот диапазон разделили на десять равных частей. Для различных символов распределения числа указанных величин в значительной степени схожи, а именно отрезки с их максимальным количеством совпали. Вне этого интервала (как слева, так и справа) имеется незначительная часть (0.1% – 0.3%) таких слагаемых, причем их количество уменьшается к периферии. Имеющееся на данном отрезке (максимальное) число рассматриваемых слагаемых, нормированное на количество изображений каждого из символов, приблизительно одинаково для различных символов. В этом интервале значения $a^{lk0}\,x_l^j$ лежат в окрестности нуля, из них чистому нулю соответствует примерно половина. Количество слагаемых, для которых $a^{lk0}\,x_l^j = 0$, совпадает с их числом при $x_l^j = 0$. При этом оказалось, что $a^{lk0} = 0$ лишь для очень незначительного числа слагаемых. Для каждого из рассматриваемых символов таковых оказалось всего лишь 16. Это свидетельствует об оптимальности полученного набора мономов, за вычетом указанного незначительного количества.

При поступлении на вход различных образов одного и того же символа с номером k_0 определим, в каком диапазоне находятся вариации составляющих $a^{lk}\,x_l^j$ и как они распределены внутри этого диапазона.

Если принимается во внимание только одна альтернатива, совпадающая с входящим символом, имеющим номер k_0, то для каждого l ($1 \le l \le L$) величина $a^{lk0}\,x_l^j$ варьируется в некотором диапазоне:

$$\sigma_l^{k0} = \max_j(a^{lk0}x_l^j) - \min_j(a^{lk0}x_l^j) \ge 0, \qquad (15)$$

где $1 \le j \le J_{k0}$.

Определим минимальную и максимальную величины такой вариации:

$$z^{k0} = \min_l(\sigma_l^{k0}) = 0, Z^{k0} = \max_l(\sigma_l^{k0}), \qquad (16)$$

Поделим полученный отрезок $[0, Z^{k0}]$ на десять равных частей и выясним, какое количество значений l имеет вариации, относящиеся к каждому такому маленькому отрезку.

В этом распределении большая доля значений l (9.0% - 99.6%) имеет незначительные вариации компонент $a^{lk0}\,x_l^j$. Отметим, что при рассмотрении всех альтернатив распознавания распределение «вытягивается» и становится более пологим.

Полученные результаты, относящиеся к распределению составляющих $a^{lk}\,x_l^j$ и их вариаций, являются обоснованием устойчивости распознавания. Во-первых, максимум обоих распределений находится в окрестности нуля, а во-вторых, в этом интервале находится подавляющее большинство соответствующих значений. Следовательно, искажение исходных изображений, заключающееся в изменении некоторого количества мономов, приводит к незначительным поправкам в оценке посредством ее составляющих $a^{lk}\,x_l^j$.

Статистический анализ функции оценки искаженных изображений

Распознавание проводилось на модификации указанной ранее базы рукопечатных цифр.

Пусть на вход распознавателя поступают перенумерованные изображения ($1 \le j \le J_{k0}$) только одного определенного k_0-го символа (цифра «8»).

Исследуем, какие изменения происходят с совокупностью значений $\{a^{lk} x_l^j\}$ при распознавании множества, элементы которого получены в результате затемнения элементов обучающего множества *(см. выше)*.

В более ранних исследованиях [3, 8] показано, что картина распознавания для различных символов не имеет существенных различий. Поэтому рассмотрен только один символ. В *таблице 1* представлено изменение количества неправильно распознанных его изображений с усилением затемнения.

Таблица 1. *Число ошибок при различном затемнении*

Степень затемнения, *n*	Число ошибок
0	138
10	71
20	106
30	339
40	2957
50	6457
60	9761
70	10116
80	10121
90	10121

Определим, в каком диапазоне лежат значения $a^{lk0} x_l^j$, а также как они распределены внутри этого диапазона для случая, когда в качестве альтернативы распознавания выбирается только сам символ. Нижняя и верхняя границы соответствующего диапазона t^{k0} и T^{k0} вычисляются по формулам:

$$t^{k0} = \min_{l,j}(a_l^{lk0} x_l^j),$$
$$T^{k0} = \max_{l,j}(a_l^{lk0} x_l^j), \qquad (17)$$
$$1 \le l \le L, 1 \le j \le J_{k0}.$$

Таблица 2. *Диапазоны значений $a^{lk0} x_l^j$ в зависимости от степени затемнения n*

Степень затемнения, *n*	Граничные значения $a^{lk0} x_l^j$	
	$\min_{l,k}(a^{lk} x_l^j)$	$\max_{l,k}(a^{lk} x_l^j)$
n = 0	-0.037	0.051
n > 0	-0.037	0.061

В *таблице 2* приведены указанные диапазоны для различных значений *n*.

Для единообразия будем рассматривать только больший диапазон [-0.037; 0.061]. Делим его на 10 равных частей и для каждого $n \ge 0$ определяем, сколько значений $a^{lk0} x_l^j$ оказывается на каждом маленьком отрезке.

На начальном этапе затемнения точность распознавания улучшается, при этом для $n = 10$ число неправильных распознаваний минимально. Такое поведение точности распознавания коррелируется с особенностью изменения картины распределения количества значений $a^{lk0} x_l^j$: на первых шагах увеличения n значение максимума количества указанных слагаемых, наблюдающегося в четвертом интервале (пик распределения в окрестности нуля (*см. выше*)), увеличивается и достигает наибольшей величины, когда $n = 30$ (в *таблице 3* обозначено жирным курсивом). Нарастание n приводит к тому, что в близлежащих интервалах (третьем, пятом и шестом) число мономов падает и достигает минимума при $n = 10$ в третьем интервале, при $n = 30$ – в пятом и $n = 20$ – в шестом (в *таблице 3* выделен жирным курсивом). При дальнейшем увеличении n картина распределения постепенно «расплывается», число слагаемых $a^{lk0} x_l^j$ в четвертом (пиковом) интервале распределения уменьшается, а во всех других, включая близлежащие, увеличивается.

Заключение

Для разработанного авторами метода распознавания символов предложена и реализована методология анализа механизма классификации. Исследована структура матрицы весовых коэффициентов a^{lk}. Проведенное на обучающем множестве рукопечатных символов большого объема статистическое исследование функции оценки позволило выявить особенности распределения числа составляющих $a^{lk} x_l^j$, входящих в виде слагаемых в формулу

DOI: 10.22204/2410-4639-2016-092-04-124-134 № 4 (92) октябрь–декабрь 2016 г.

Таблица 3. *Распределение количества значений $a^{lk0}\,x_l^j$ в интервалах диапазона [-0.037; 0.061]*

Степень затемнения, *n*	Количество значений $a^{lk0}\,x_l^j$ в интервалах диапазона [-0.037; 0.061]									
	1	2	3	4	5	6	7	8	9	10
0	124	659	126309	221647273	2667904	49653	1926	40	13	0
10	215	1094	***107223***	221802217	2531081	49079	2928	46	14	3
20	343	1564	111413	221931741	2397075	***47650***	4020	61	22	7
30	562	2281	124882	***222059446***	***2253006***	48436	5121	103	33	22
40	799	3395	176639	221622472	2623139	59349	7210	620	217	34
50	1105	5028	227901	221269448	2899881	80067	9264	816	291	57
60	1523	7829	311359	220687930	3357950	112792	12255	1644	425	113
70	2122	12580	457786	219748030	4067998	184364	17054	2233	1352	237
80	3419	24895	758029	218359711	5030651	278477	30515	4761	2672	478
90	6285	74769	1313762	216762545	5769642	479803	75670	2117	5640	3115

для ее вычисления. Подтверждена оптимальность используемого набора мономов. Характер распределения числа составляющих $a^{lk}\,x_l^j$ и их вариации является обоснованием устойчивости распознавания. Аналогичные исследования проведены на множествах, полученных в результате искажения исходного, причем величина различия между ними постепенно увеличивается. Показано, что поведение точности распознавания коррелируется с характером изменения картины распределения числа слагаемых $a^{lk}\,x_l^j$.

Литература

1. ***А.В. Мисюрев***
   В *Интеллектуальные технологии ввода и обработки информации. Серия: Труды ИСА РАН,* под ред. В.Л. Арлазарова, Н.Е. Емельянова, Москва, URSS. 1998, 164 с.
2. ***М.Б. Гавриков, А.В. Мисюрев, Н.В. Пестрякова, О.А. Славин***
   Автоматика и Телемеханика, 2006, №2, 119.
3. ***Н.В. Пестрякова***
   Информ. технол. вычисл. сист., 2010, №2, 75.
4. ***М.Б. Гавриков, Н.В. Пестрякова***
   Метод полиномиальной регрессии в задачах распознавания печатных и рукопечатных символов, Препринты ИПМ им. М.В. Келдыша, 2004, 022, 12 с. (http://www.keldysh.ru/papers/2004/prep22/prep2004_22.html).
5. ***Ю.В. Линник***
   Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений, Москва, Физматлит, 1958, 336 с.
6. ***О.В. Локуциевский, М.Б. Гавриков***
   Начала численного анализа, Москва, ТОО «Янус», 1995, 582 с.
7. ***J. Schürmann***
   Pattern Classification: A Unified View of Statistical and Neural Approaches, USA, NY, New-York, Wiley-Interscience Publ., 1996, 392 с.
8. ***Н.В. Пестрякова***
   Информ. технол. вычисл. сист., 2009, №1, 58.

English

Multivariate Numerical Modelling in Solving the Research Problem of Statistical Pattern Recognition Methods' Stability to Distortion of Images*

Boris M. Gavrikov –
Institute for System Analysis FRC
"Computer Science and Control" RAS
9, 60-letiya Octyabrya Ave.,
Moscow, 117312, Russia
e-mail: bmgavrikov@gmail.com

Nadezhda V. Pestryakova –
Institute for System Analysis FRC
"Computer Science and Control" RAS
9, 60-letiya Octyabrya Ave.,
Moscow, 117312, Russia
e-mail: pestryakova@isa.ru

Abstract

The paper describes the probabilistic way of classification based on polynomial regression approach. This method is used for recognition of printed and hand-printed symbols. The method stability to images distortion is proved by means of multivariate numerical modelling.

Keywords: classification, polynomial regression, printed and hand-printed symbols.

Images & Tables

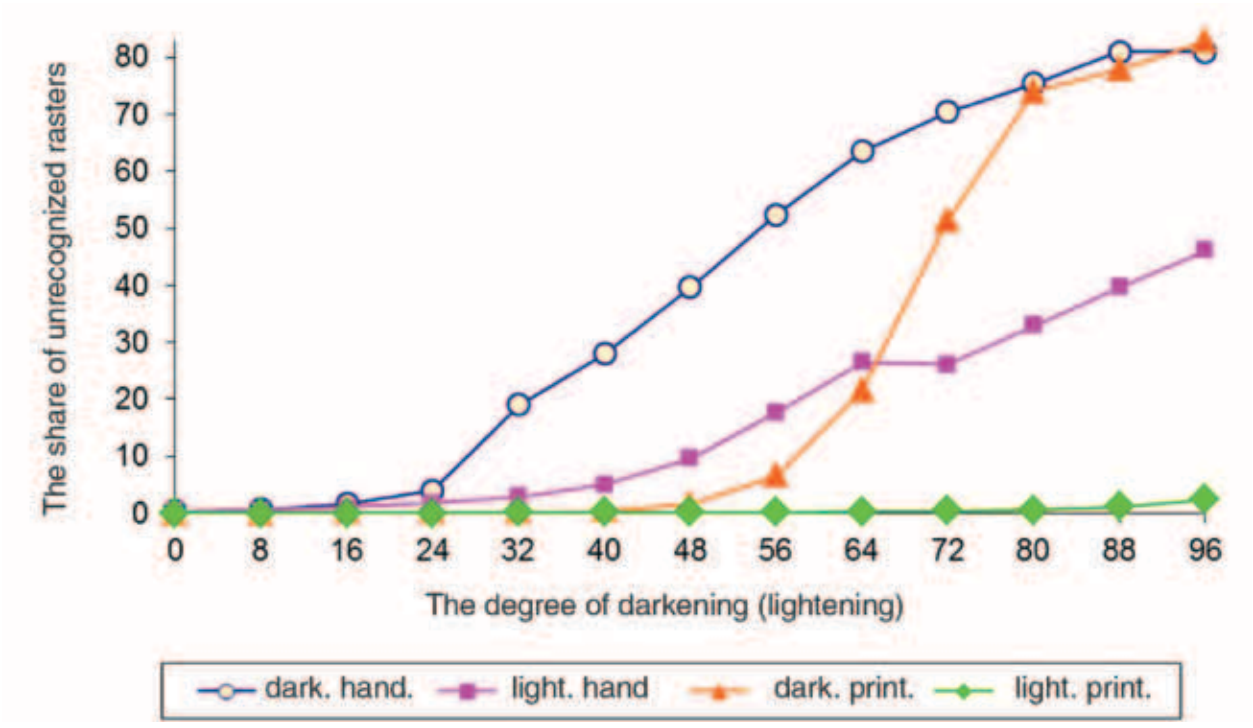


Fig. 1. The share of unrecognized rasters provided darkening or lightening.

* *The work was financially supported by RFBR (projects 13-07-00262 and 13-07-12176).*

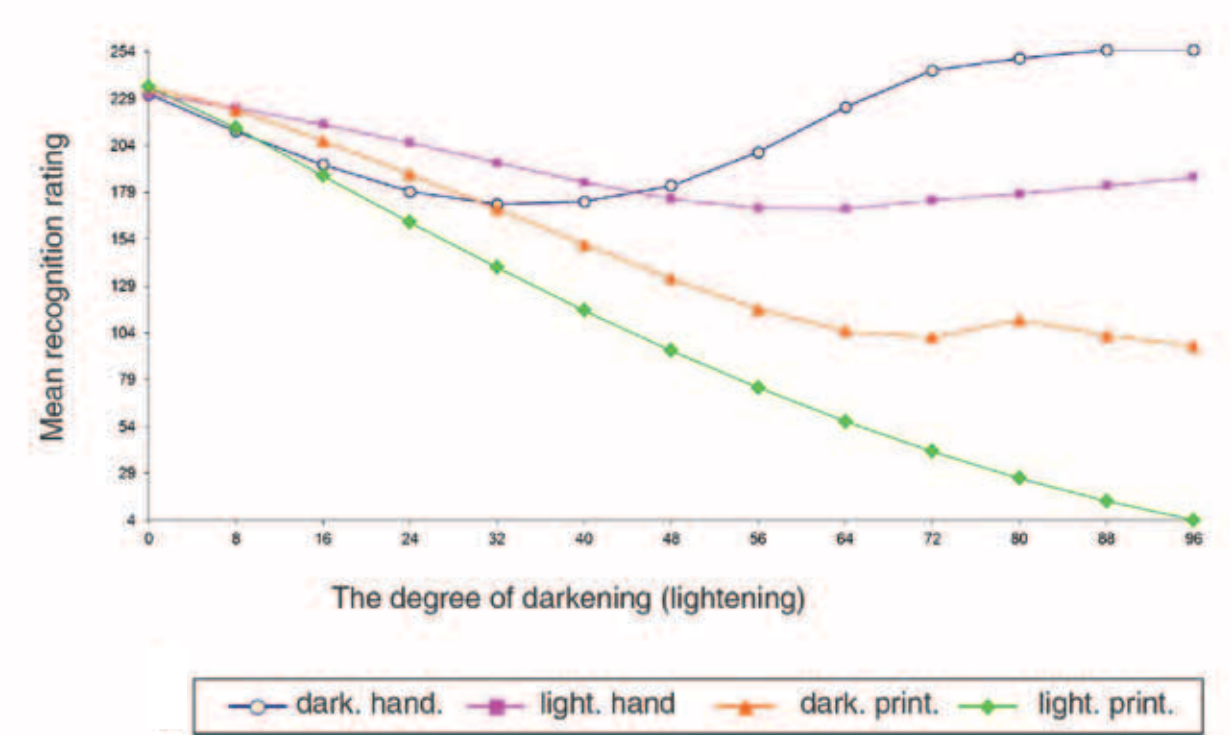


Fig. 2. Mean recognition rating provided darkening or lightening.

Table 1. *The number of errors at different degree of darkening*

Degree of darkening, *n*	Number of errors
0	138
10	71
20	106
30	339
40	2957
50	6457
60	9761
70	10116
80	10121
90	10121

Table 2. *The ranges of* $a^{lk0}\, x_l^j$ *values depending on the n degree of darkening*

Degree of darkening, *n*	Boundary values $a^{lk0}\, x_l^j$	
	$\min_{l,k}(a^{lk} x_l^j)$	$\max_{l,k}(a^{lk} x_l^j)$
$n = 0$	-0.037	0.051
$n > 0$	-0.037	0.061

Table 3. *The distribution of the number of* $a^{lk0}\,x_l^j$ *values in the intervals of the range [-0.037; 0.061]*

Degree of darkening, *n*	Number of $a^{lk0}\,x_l^j$ values in the intervals of the range [-0.037; 0.061]									
	1	2	3	4	5	6	7	8	9	10
0	124	659	126309	221647273	2667904	49653	1926	40	13	0
10	215	1094	***107223***	221802217	2531081	49079	2928	46	14	3
20	343	1564	111413	221931741	2397075	***47650***	4020	61	22	7
30	562	2281	124882	***222059446***	***2253006***	48436	5121	103	33	22
40	799	3395	176639	221622472	2623139	59349	7210	620	217	34
50	1105	5028	227901	221269448	2899881	80067	9264	816	291	57
60	1523	7829	311359	220687930	3357950	112792	12255	1644	425	113
70	2122	12580	457786	219748030	4067998	184364	17054	2233	1352	237
80	3419	24895	758029	218359711	5030651	278477	30515	4761	2672	478
90	6285	74769	1313762	216762545	5769642	479803	75670	2117	5640	3115

References

1. ***A.V. Misyurev***
   In *Intellectual Technologies of Data Entry and Processing. Ser. Proc. of Institute for System Analysis RAS [Intellektualnye tekhnologii vvoda i obrabotki informatsii. Ser. Trudy ISA RAN],* Eds V.L. Arlazarov, N.E. Emelyanov, Moscow, URSS Publ., 1998, 164 pp. (in Russian).
2. ***M.B. Gavrikov, A.V. Misyurev, N.V. Pestryakova, O.A. Slavin***
   Automation and Remote Control, 2006, **67**(2), 278. DOI: 10.1134/S000511790602007X.
3. ***N.V. Pestryakova***
   J. Information Technology and Computer Systems [Informatsionnye tekhnologii i vychislitelnye sistemy], 2010, №2, 75 (in Russian).
4. ***M.B. Gavrikov, N.V. Pestryakova***
   Polynomial Regression in Recognition of Printed and Hand-Printed Letters, Preprint, M. Keldysh Inst. Appl. Math. RAS, 2004, 022, 12 pp. (in Russian). (http://www.keldysh.ru/papers/2004/prep22/prep2004_22.html).
5. ***Yu.V. Linnik***
   Method of Least Squares and Principles of the Theory of Observations, USA, NY, New York, Pergamon Press, 1961, 360 pp.
6. ***O.V. Lokutsievskiy, M.B. Gavrikov***
   The Basics of Numerical Analysis [Nachala chislennogo analiza], Moscow, Yanus Publ., 1995, 582 pp. (in Russian).
7. ***J. Schürmann***
   Pattern Classification: A Unified View of Statistical and Neural Approaches, USA, NY, New-York, Wiley-Interscience Publ., 1996, 392 c.
8. ***N.V. Pestryakova***
   J. Information Technology and Computer Systems [Informatsionnye tekhnologii i vychislitelnye sistemy], 2009, №1, 58 (in Russian).

Исследование критериев оценки научных проектов с помощью методов машинного обучения на примере конкурсов РФФИ *

Д.А. Девяткин, Р.Е. Суворов, И.А.Тихомиров, О.Г. Григорьев

В работе представлены результаты анализа основных критериев оценки научных проектов, используемых научными фондами. Приведен краткий обзор методов анализа экспертных анкет при принятии решений о финансировании проектов. Предложен новый подход к определению значимости критериев оценки научных проектов, основанный на методах машинного обучения. Этот подход позволяет работать с качественными и количественными критериями на больших объемах данных и не требует приведения значений критериев к численному виду. Проведены экспериментальные исследования по определению значимости критериев оценки научных проектов на примере инициативных конкурсов РФФИ. Показано, что состав наиболее значимых критериев оценки проектов остается практически неизменным для всех научных областей, за исключением направления «Естественнонаучные методы исследования в гуманитарных науках». Сделаны выводы о возможности использования предложенного подхода для верификации итоговых оценок экспертов, а также для проверки значимости вновь вводимых критериев.

Ключевые слова: оценка научных проектов, значимость критериев, случайный лес (random forest), машинное обучение.

** Работа выполнена при финансовой поддержке РФФИ (проект № 14-29-05075).*

Введение

Поддержка коллективов и отдельных ученых в форме финансирования исследовательских проектов является важным инструментом управления фундаментальной наукой [1]. В мире существует большое количество научных фондов, распределение финансирования в них проводится на конкурсной основе: каждая заявка анализируется несколькими экспертами, которые заполняют экспертные анкеты. Затем на основании этих анкет принимается решение о выделении финансирования либо об отклонении заявки. Каждая анкета содержит фиксированный набор критериев, которые определяются политикой фонда *(табл. 1)*.

Известно, что значимость критериев в экспертной анкете, то есть степень их влияния на результат экспертизы, неодинакова и обычно оценивается группами экспертов при помощи достаточно трудоемких эвристических подходов [10]. В то же время многие научные фонды, в частности РФФИ, за время своей деятельности накопили крупные массивы заявок, прошедших экспертизу, и экспертных анкет, что позволяет использовать методы машинного обучения для определения относительной значимости критериев оценки научных проектов. Такой подход может применяться как для ретроспективного анализа существующих анкет фондов с целью исключения незначимых критериев и снижения трудоемкости экспертизы проектов, так и для определения состоятельности новых критериев, предлагаемых научным сообществом [11, 12].

Целью настоящего исследования является разработка метода определения значимости качественных и количественных критериев оценки научных проектов, который позволяет работать с большими объемами данных, и его экспериментальная оценка.

ДЕВЯТКИН
Дмитрий Алексеевич
Институт системного анализа ФИЦ «Информатика и управление» РАН

СУВОРОВ
Роман Евгеньевич
Институт системного анализа ФИЦ «Информатика и управление» РАН

ТИХОМИРОВ
Илья Александрович
Институт системного анализа ФИЦ «Информатика и управление» РАН

ГРИГОРЬЕВ
Олег Георгиевич
Институт системного анализа ФИЦ «Информатика и управление» РАН

Таблица 1. *Основные критерии оценки проектов в различных научных фондах*

Страна	Организация/ Фонд	Бюджет (млрд руб.)*	Год	Основные критерии
Великобритания	EPSRC [2]	80.2	2009	1. Качество исследования, проводимого заявителем. 2. Степень значимости для государства. 3. Влияние данного гранта на связанные проекты. 4. Полнота обзора связанных работ. 5. Проработанность стратегии управления проектом. 6. Степень эффективности запланированных исследований – заявитель должен обосновать эффективность вложения средств. 7. Предусмотрены ли в рамках проекта мероприятия, направленные на популяризацию технических и физико-математических наук.
Германия	DFG [3]	143.6	2013	1. Качество и оригинальность исследования. 2. Уровень квалификации вовлеченных специалистов. 3. Наличие четкой формулировки проверяемой гипотезы и проработанной стратегии выполнения проекта. 4. Наличие доступной технологической базы для проведения исследований.
США	NSF [4]	435.7	2012	1. Оценка научного потенциала проекта: может ли реализация проекта привести к расширению научного знания. 2. Социальная значимость проекта.
	NIH [5]	2026.9	2010	1. Научная значимость результатов. 2. Уровень квалификации вовлеченных специалистов. 3. Потенциал внедрения результатов в клиническую практику. 4. Проработанность предлагаемого подхода: позволят ли предлагаемые методы и стратегии добиться декларируемых результатов. 5. Наличие инфраструктуры, необходимой для проведения проекта.
ЕС (головной офис – Франция)	ESF [6]	3.6	2010	1. Релевантность тематике конкурса. 2. Оценка научного уровня заявки. 3. Социальная значимость проекта. 4. Оценка уровня руководителя проекта и проработанности стратегии управления. 5. Влияние данного гранта на связанные проекты. 6. Уровень взаимодействия с другими исследовательскими группами. 7. Оценка возможного синергетического эффекта, вызванного реализацией проекта.
Россия	РФФИ [7]	9.2	2014	1. Оценка предыдущих научных результатов авторов (вне связи с рассматриваемым проектом). 2. Общая оценка уровня фундаментальности проекта. 3. Обоснованность методов решения проблемы. 4. Реализуемость проекта. 5. Формулировка проблемы и целей исследования. 6. Степень значимости ожидаемых результатов. 7. Характер предполагаемых исследований.
	РНФ [8]	11.4	2014	1. Соответствие тематики проекта научным направлениям, поддерживаемым РНФ. 2. Профессиональный уровень руководителя проекта и научного коллектива. 3. Научная обоснованность проекта. 4. Значимость результатов выполнения проекта. 5. Качество планирования проекта.
Финляндия	Suomen Akatemia [9]	24.6	2011	1. Научное качество и инновационность плана исследований. 2. Оценка компетенции исполнителей. 3. Реализуемость плана исследований. 4. Наличие контактов с другими исследовательскими коллективами. 5. Значение проекта для дальнейшего профессионального роста исследователей.

* По обменному курсу ЦБ РФ на 24.11.2015.

 DOI: 10.22204/2410-4639-2016-092-04-135-146 № 4 (92) октябрь–декабрь 2016 г.

Таблица 2. *Результаты экспериментов по классификации проектов с учетом специфики научных областей*

Научная область	Количество примеров	Количество принятых заявок	Количество отклоненных заявок	F_1, %	P, %	R, %
Математика, механика и информатика	1859	581	1278	74.1	80.8	68.6
Физика и астрономия	2749	841	1908	67.0	80.0	57.6
Химия	2486	1734	752	69.5	82.7	60.2
Биология и медицинские науки	4097	1265	2832	74.3	79.9	69.8
Науки о Земле	2137	1479	658	66.1	78.3	57.7
Естественнонаучные методы исследования в гуманитарных науках	1367	318	1049	69.8	71.3	84.9
Инфокоммуникационные технологии и вычислительные системы	1988	589	1399	76.1	82.9	70.8
Фундаментальные основы инженерных наук	2738	828	1910	69.4	79.5	62.6

Связанные работы

Во многих работах отмечается, что подходы, использующие исключительно количественные критерии, плохо подходят для экспертизы научных проектов [13, 14]. Из-за этого научные фонды зачастую используют качественные критерии оценки (*табл. 1 и 2*), прямое превращение которых в численные показатели затруднено или может привести к заведомо некорректным результатам оценки проекта [15].

Вместе с тем предполагается [14, 16], что применение количественных критериев, в том числе наукометрических, может способствовать повышению качества экспертизы. Поэтому для анализа значимости критериев необходимо использовать такие методы машинного обучения, которые позволяют работать с качественными и количественными оценками, значения которых получены от нескольких экспертов.

Среди опубликованных результатов исследований тема определения значимости критериев оценки научных проектов затрагивается в работах, посвященных методам принятия решений о поддержке научно-исследовательской работы (НИР). Например, в работе [17] предложен автоматизированный метод поддержки принятия решений, в основе которого лежит метод метрической классификации объектов, для их представления используются мультимножества. Элементами мультимножества при этом являются наборы оценок заявки, данные несколькими экспертами. Другой подход к решению этой задачи [12]: для принятия решений о поддержке научного проекта на основе неполных данных предложена модель оценки проекта, в которой пропущенные оценки заполняются оценками «похожих» экспертов. Для поиска таких экспертов используется мера расстояния очевидности [18].

Следует отметить также несколько работ, посвященных методам формирования интегральных оценок научных проектов и заявок на проведение НИР, например метод формирования интегральной оценки уровня проекта на основе числовых и категориальных оценок нескольких экспертов [15]. Для вычисления этой оценки использовался метод вербального анализа решений, позволяющий преобразовывать множество входных критериев в небольшое число составных признаков.

В многокритериальном подходе [19] к оценке заявок на проведение НИР, основанном на использовании теории интервальных интуиционистских нечетких множеств [20], для определения относительной значимости критериев в условиях недостаточной информации строится интервальная интуиционистская матрица нечетких предпочтений экспертов.

В приведенных работах, однако, не рассматривается определение значимости критериев оценки на основе больших массивов данных. Тем не менее существует класс задач в области машинного обучения, близких к изучаемой в этой статье проблеме – из-

влечение ключевых признаков (feature selection). Достаточно известным подходом [21], применяемым для решения таких задач, является использование «случайного леса» (random forest). В работе [22] экспериментально доказывается устойчивость такого подхода к зашумленным данным по сравнению с методами SVM и RELIEFF. Обучение классификатора в случайном лесе [23] организовано таким образом, чтобы каждое дерево обучалось независимо на случайно выбранном подмножестве обучающих примеров. При этом при добавлении очередного узла в дерево учитывается случайно выбранное подмножество признаков. Благодаря тому, что все деревья в композиции строятся независимо, метод хорошо поддается распараллеливанию и подходит для обработки значительных объемов данных. После обучения для каждого дерева, входящего в композицию, вычисляются оценки значимости соответствующих ему признаков.

Исходя из цели данной работы, для определения значимости критериев оценки научных проектов был выбран случайный лес, который позволяет получить интерпретируемые и устойчивые к зашумленным данным результаты обучения и подходит для работы с большими наборами данных.

Методика проведения эксперимента

Описание массива тестовых данных. В тестовый массив были включены данные форм заявок и экспертных анкет по проектам РФФИ «А» за 2013–2014 гг., а также информация о поддержке этих проектов фондом, всего 19421 проект. Указанный интервал дат был выбран по причине того, что в этот период состав критериев оценки проектов не изменялся. К каждой заявке в тестовом массиве прилагается также по три заполненные экспертные анкеты. Более 65% массива составили отклоненные проекты. В качестве исследуемых критериев выступали все поля экспертных анкет и заявок: текстовые, количественные и категориальные (принимают значения из конечного множества, причем на элементах этого множества не установлен порядок).

Случайный лес позволяет работать исключительно с числовыми данными, поэтому выполнялось преобразование значений категориальных и текстовых критериев в наборы количественных признаков. Для представления текстовых критериев в виде набора количественных признаков использовали модель «мешок слов». Перевод значений категориальных критериев в числовую форму проводили путем бинаризации. Таким образом, в качестве итогового пространства признаков C использовали объединение множества числовых критериев NR, текстовых признаков TR и декартового произведения множества всех категориальных критериев $Cr \in CR$ и множества их значений K_{Cr}:

$$C = \bigcup_{\forall Cr \in CR} (Cr \times K_{Cr}) \bigcup NR \bigcup TR ,$$

Далее, говоря о признаках, мы будем иметь в виду объекты из множества C, а под критериями будем понимать поля экспертных анкет и заявок. Всего на основе форм заявок и экспертных анкет было сгенерировано более 5 тыс. признаков.

Описание хода экспериментов. В рамках первого эксперимента оценивали потенциальную пригодность критериев, используемых в ходе экспертизы проектов РФФИ, для автоматического анализа проектов с помощью методов машинного обучения. Для этого при помощи случайного леса классифицировали тестовой массив данных по двум классам: «поддержанные проекты» и «отклоненные проекты». Этот метод представляет собой композицию классификаторов на деревьях решений (лес) $H = \{h(x,\Theta_k), k,= 1, ...\}$, где $\{\Theta_k\}$ – векторы параметров, настраиваемых при обучении, и каждое дерево участвует в голосовании за наиболее популярный класс, соответствующий вектору признаков x. Для выбора разделяющего признака используется критерий неоднородности по Джини [24]. Оптимальные гиперпараметры для работы классификатора подбирали методом случайного поиска [25]. Для получения эмпирических оценок качества классификации использовали метод скользящего контроля по трем блокам с макроусреднением. Качество классификации оценивалось с помощью показателей полноты, точности и F_1-меры. Для того чтобы оценить вклад отдельных групп критериев, таких как оценки экспертных анкет, итоговые оценки, а также поля, взятые из форм заявок, были проведе-

 DOI: 10.22204/2410-4639-2016-092-04-135-146 № 4 (92) октябрь–декабрь 2016 г.

ны дополнительные эксперименты по классификации проектов на подмножествах признаков.

Во втором эксперименте оценивали достаточность критериев, применяемых в ходе экспертизы проектов, для анализа с помощью методов машинного обучения. Для этого классифицировали проекты случайным лесом на наборе данных, использованных для обучения. В этом эксперименте значение F_1, близкое к единице, указывает на то, что внесение дополнительных критериев не имеет смысла, в ином случае для увеличения качества классификации могут быть введены дополнительные критерии оценки проектов. Было выявлено также, по каким именно научным областям можно улучшить качество анализа с помощью введения дополнительных критериев оценки.

В ходе третьего эксперимента проверяли, насколько хорошо средняя итоговая оценка экспертов может быть промоделирована с использованием всех остальных критериев из экспертных анкет. Предварительные эксперименты показали линейный характер зависимости итоговой оценки от других критериев, поэтому для моделирования использовали линейную регрессию $y = x^T w + \varepsilon$, где w – параметры регрессии, y – итоговая оценка, x – вектор признаков, ε – случайная ошибка, такая, что ее математическое ожидание $E(\varepsilon)=0$. Для оценки качества регрессионной модели использовали нормированный коэффициент детерминации и среднюю абсолютную ошибку. В этом эксперименте оценивали также влияние различных критериев на величину моделируемой итоговой оценки. Так как все признаки, используемые для построения регрессии, были одинаково нормированы, для определения значимости признаков применяли соответствующие элементы вектора параметров регрессии w.

В последнем эксперименте определяли значимость критериев оценки проектов (за исключением итоговой оценки) в различных научных областях. Для этого исходный тестовый набор данных был разделен на несколько поднаборов, в каждом из которых содержались заявки только по одной области науки. В качестве меры важности признаков использовали оценки значимости по Джини [24], рассмотрим способ их вычисления. Пусть $T = \{(y_i x_i), i = 1...n\}$ – множество обучающих примеров для построения леса H, а $C = \{c: c = 1...|x|\}$ – множество идентификаторов признаков для обучения. Множество $T_k \subset T$ – набор обучающих примеров для дерева $h(x, \Theta_k)$ из леса H, тогда множество $T_{obk} = T \backslash T_k$ – набор примеров для построения оценок значимости признаков на данных, не использованных при обучении (out-of-bag) для этого дерева $h(x, \Theta_k)$. Тогда $f \subset T_{obk}$ – набор примеров, соответствующий некоторому узлу m дерева $h(x, \Theta_k)$ при вычислении оценок значимости признаков. Для каждого узла решающих деревьев вычисляется неоднородность по Джини [24] I_G – мера того, как часто случайно выбранный элемент из множества данных, разделяемых в некоторой вершине, может быть некорректно классифицирован, если он был случайно размечен в соответствии с распределением классов в этом множестве. Пусть c – признак, на основании которого происходит разделение данных в вершине m на подмножества, соответствующие правой и левой дочерним вершинам R и L. Вероятности отнесения данных из вершины m к левому и правому подмножествам составляют соответственно P_L и P_R:

$$f \xrightarrow[c]{} (f_R, f_L) : f_R \cup f_L = f, f_R \cap f_L = \varnothing,$$

$$P_L = \frac{|f_L|}{|f|}, P_R = 1 - P_L \quad .$$

Здесь символом $\xrightarrow[c]{}$ обозначена операция разделения множества на основании значения признака c. Тогда снижение неоднородности по Джини ΔI_G для дочерних узлов вершины m на основании признака c составит $\Delta I_G(f, c) = I_G(f) - [P_R I_G(f_R) + P_L I_G(f_L)]$.

Значимость признака по Джини $\Delta I_{GS}(c)$ вычисляется как суммарное снижение неоднородности по Джини, вызываемое этим признаком во всех деревьях леса. Тогда значимость признака $W(c)$ определяется как суммарная значимость по Джини этого признака по итогам процедуры скользящего контроля. В работах [26, 27] и др. было эмпирически показано соответствие полученной таким образом меры значимости признака и степени его влияния на исход классификации. Затем вычисляется значимость исходных критериев I_{mp} как средняя значимость всех признаков, построенных на основе этого критерия:

$$I_{mp}(Cr) = \frac{1}{|Cr|} \sum_{\forall c \in Cr} W(c).$$

После вычисления значимости, выполняли переупорядочивание критериев:

$$\forall i, j = 1..|CR|;$$

$$i > j : I_{mp}(Cr_i) > I_{mp}(Cr_j).$$

При помощи такого подхода для каждого научного направления извлекали по 12 наиболее значимых критериев.

Для проверки универсальности используемых критериев оценки проектов вычисляли меру близости полученных упорядоченных списков критериев друг относительно друга. Состав наиболее значимых критериев может отличаться для различных научных областей, что делает невозможным применение для решения этой задачи коэффициента ранговой корреляции Спирмена: при рассмотрении объединения критериев из двух сравниваемых списков значимость и ранг некоторых элементов в одном из списков могут оказаться неопределенными. Поэтому в качестве меры близости использовали смещенную оценку перекрытия рангов [28]:

$$RBO_{ext}(S,T,p) = \frac{X(S,T,k)}{k} p^k + \frac{1-p}{p} \sum_{d=1}^{k} \frac{X(S,T,d)}{d} p^d$$

где S и T – сравниваемые упорядоченные наборы критериев, $X(S,T,d) = |S_{:d} \cap T_{:d}|$ – размер пересечения двух подмножеств $S_{:d}$ и $T_{:d}$ длины d. Эта оценка является нормированной, она изменяется в пределах от 0 до 1, где 0 соответствует двум полностью несовпадающим упорядоченным множествам, а 1 свидетельствует о полном совпадении.

Результаты экспериментов

В первом эксперименте с помощью случайного леса выполняли классификацию всех проектов «А» РФФИ за 2013–2014 гг. При этом изучалось влияние различных групп критериев на результат. Была получена оценка F_1-меры 84.7% при среднеквадратичном отклонении не более 3%. Выявлено, что наибольшее влияние на результат классификации оказывают итоговые оценки экспертов, остальные критерии также позволяют выполнять классификацию проектов, но с несколько меньшим значением F_1-меры (74.7%). Использование же только критериев, полученных из форм заявок, не позволяет получить приемлемое качество классификации.

В результате второго эксперимента на наборе данных для обучения F_1-мера составила ~88%, что лишь на 4% выше, чем оценка, полученная методом скользящего контроля. То есть для дальнейшего повышения F_1-меры необходимо расширить набор критериев, например наукометрическими показателями состоятельности исполнителей проекта, что является направлением дальнейших исследований. В этом эксперименте классификатор был отдельно обучен и протестирован на каждой научной области. Области, получившие относительно низкую оценку по F_1-мере, вероятно, в большей степени зависят от дополнительных критериев оценки, не учтенных при проведении эксперимента *(табл. 2)*.

В ходе третьего эксперимента значения итоговой оценки проекта моделировали при помощи линейной регрессии на основании всех остальных критериев, взятых из экспертных анкет. Метод показал хорошие результаты работы: средняя абсолютная ошибка составила 0.45, притом что сама моделируемая величина может изменяться в интервале от 1 до 9, а нормированный коэффициент детерминации превысил 0.83. Таким образом, значения критерия «итоговая оценка» связаны со значениями остальных критериев зависимостью, близкой к функциональной. Такую регрессионную модель можно использовать для верификации итоговых оценок экспертов, информируя о случаях большого расхождения величины итоговой оценки со значениями других критериев. Получен также список критериев, наиболее сильно влияющих на величину итоговой оценки. К таким критериям относятся «Реализуемость проекта», «Общая оценка уровня фундаментальности проекта», «Степень значимости ожидаемых результатов» и «Оценка предыдущих научных результатов авторов».

В ходе четвертого эксперимента определяли значимость критериев оценки проектов в зависимости от научной области. Итоговые оценки экспертов при этом не учитывались.

DOI: 10.22204/2410-4639-2016-092-04-135-146 № 4 (92) октябрь–декабрь 2016 г.

Таблица 3. *Наиболее важные критерии оценки проектов с учетом специфики научных областей*

Критерий	Ранг критерия в научной области							
	Математика, механика и информатика	Физика и астрономия	Химия	Биология и медицинские науки	Науки о Земле	Естественнонаучные методы исследования в гуманитарных науках	Инфокоммуникационные технологии и вычислительные системы	Фундаментальные основы инженерных наук
Оценка предыдущих научных результатов авторов	1	1	2	1	3	7	1	1
Общая оценка уровня фундаментальности проекта	2	4	6	6	6	1	3	4
Обоснованность методов решения проблемы	3	3	1	2	1	4	5	2
Формулировка проблемы и целей исследования	4	5	4	5	4	2	6	6
Степень значимости ожидаемых результатов	5	2	3	4	5	5	4	3
Число публикаций руководителя	6	7	7	9	8	-*	10	9
Ученые степени участников	7	-	-	-	12	-	-	-
Реализуемость проекта	8	6	5	3	2	3	2	5
Количество членов научного коллектива	9	12	8	-	-	12	7	-
Возраст руководителя	10	10	12	10	11	10	12	11
Ученые звания участников	11	-	9	-	-	9	8	-
Количество публикаций участников	12	11	11	12	9	-	-	10
Возраст участников	-	8	-	11	7	11	11	8
Число основных публикаций коллектива, наиболее близко относящихся к предлагаемому проекту	-	9	-	7	10	-	-	7
Запрашиваемый объем финансирования	-	-	10	8	-	-	9	12
Научная дисциплина	-	-	-	-	-	6	-	-
Должность участника проекта	-	-	-	-	-	8	-	-

* Критерий не является значимым для соответствующей области науки.

Ранжированные списки наиболее значимых критериев оценки проектов представлены в *таблице 3*. В ней выделены критерии, вошедшие в топ-5 по важности для всех научных направлений. На первом месте по важности оказались качественные критерии, взятые из экспертных анкет.

Между полученными ранжированными списками была вычислена мера близости RBO_{12} при значении скорости убывания значимости критериев $p = 0.7$ *(табл. 4)*. Эта мера близости оказалась достаточно высокой для всех научных отраслей, кроме области, посвященной естественнонаучным методам исследования в гуманитарных науках. Это означает, что состав наиболее значимых критериев, используемых для оценки проектов, в целом слабо зависит от научной области.

Согласно результатам экспериментов, наиболее значимыми оказались критерии, взятые из экспертных анкет, такие как оценка предыдущих научных результатов авторов, общая оценка уровня фундаментальности проекта, степень значимости ожидаемых результатов и обоснованность методов решения проблемы. Выявлено, что состав критериев, наиболее сильно влияющих на результат экспертизы проектов, выполняемых в рамках инициативных конкурсов РФФИ, практически не зависит от научной области, за исключением области «Естественнонаучные методы исследования в гуманитарных науках».

Таблица 4. *Близость наиболее значимых критериев оценки фундаментальных проектов по* RBO_{12}

Научная область	Математика, механика и информатика	Физика и астрономия	Химия	Биология и медицинские науки	Науки о Земле	Естественнонаучные методы исследования в гуманитарных науках	Инфокоммуникационные технологии и вычислительные системы	Фундаментальные основы инженерных наук
Математика, механика и информатика	1.00	0.79	0.48	0.74	0.37	0.41	0.74	0.77
Физика и астрономия	0.79	1.00	0.54	0.79	0.37	0.24	0.73	0.87
Химия	0.48	0.54	1.00	0.60	0.80	0.24	0.42	0.64
Биология и медицинские науки	0.74	0.79	0.60	1.00	0.55	0.29	0.78	0.90
Науки о Земле	0.37	0.37	0.80	0.55	1.00	0.31	0.45	0.47
Естественнонаучные методы исследования в гуманитарных науках	0.41	0.24	0.24	0.29	0.31	1.00	0.34	0.24
Инфокоммуникационные технологии и вычислительные системы	0.74	0.73	0.42	0.78	0.45	0.34	1.00	0.75
Фундаментальные основы инженерных наук	0.77	0.87	0.64	0.90	0.47	0.24	0.75	1.00

Заключение

В работе предложен новый подход к определению значимости критериев оценки научных проектов, который позволяет работать с качественными и количественными критериями на больших объемах данных.

В результате экспериментов на тестовом наборе данных выделены наиболее значимые критерии оценки проектов. Состав этих критериев оказался практически неизменным для различных научных областей, за исключением области, посвященной естественнонаучным методам исследования в гуманитарных науках. Выявлено, что наиболее значимыми оказались качественные критерии, полученные из экспертных анкет. Обнаружена зависимость, близкая к функциональной, между значением критерия «Итоговая оценка» и значениями остальных критериев из экспертных анкет. Эта зависимость может использоваться для верификации итоговых оценок экспертов, например, если проект при сходных значениях критериев получает сильно отличающееся значение итоговой оценки по сравнению с другими заявками (заниженное или завышенное). Кроме того, предложенный подход можно использовать для проверки значимости вновь вводимых критериев и оценки их влияния на принятие/отклонение проектов. В целом предлагаемый подход может повысить качество и объективность экспертизы научных проектов, а также снизить нагрузку на грантозаявителей и экспертов путем удаления незначимых полей из заявок и экспертных анкет.

В дальнейших исследованиях планируется провести эксперименты по расширению набора критериев для оценки заявок дополнительными показателями, которые можно получать автоматизированно при помощи систем поддержки научно-технической деятельности [29]. Также планируется расширить предлагаемый подход: учесть различия в оценке фундаментальных и прикладных исследований, анализировать текст обоснования оценок на предмет их «прозрачности» и провести аналогичные эксперименты на данных других научных фондов и организаций, финансирующих научные проекты. Кроме того, будут проведены аналогичные эксперименты с использованием альтернативных случайному лесу машинных методов и их ансамблей.

Литература

1. ***M. Benner, U. Sandström***
Research Policy, 2000, **29**(2), 291. DOI: 10.1016/S0048-7333(99)00067-0.
2. *EPSRC – Engineering and Physical Sciences Research Council.* (https://www.epsrc.ac.uk/).
3. *DFG – Deutsche Forschungsgemeinschaft.* (http://www.dfg.de/en/).
4. *NSF – National Science Foundation.* (https://www.nsf.gov./).
5. *NIH – National Institutes of Health.* (https://www.nih.gov/).
6. *ESF – European Science Foundation.* (http://www.esf.org/).
7. *РФФИ – Российский фонд фундаментальных исследований.* (http://www.rfbr.ru/rffi/ru/).
8. РНФ – *Российский научный фонд.* (http://grant.rscf.ru/).
9. *Academy of Finland.* (http://www.aka.fi/en/).
10. ***K. Olson***
Field Methods, 2010, **22**(4), 295. DOI: 10.1177/1525822X10379795.
11. ***A. Geuna, B.R. Martin***
Minerva, 2003, **41**(4), 277.
DOI: 10.1023/B:MINE.0000005155.70870.bd.
12. ***J. Zhu, H. Wang, C. Ye, Q. Lang***
Information Fusion, 2015, **24**(C), 93.
DOI: 10.1016/j.inffus.2014.09.006.
13. ***О.И. Ларичев***
Наука и искусство принятия решений, Москва, Изд. Наука, 1979, 200 с.
14. ***P. Weingart***
Scientometrics, 2005, **62**(1), 117. DOI: 10.1007/s11192-005-0007-7.
15. ***А.Б. Петровский, Г.В. Ройзензон, А.В. Балышев, И.П. Тихонов***
IJ IMA, 2012, **1**(4), 349.
16. ***S. Wessely***
The Lancet, 1998, **352**(9124), 301. DOI: 10.1016/S0140-6736(97)11129-1.
17. ***А.Б. Петровский***
Информационные технологии и вычислительные системы, 2004, №2, 56.
18. ***A.-L. Jousselme, D. Grenier, E. Bossé***
Information Fusion, 2001, **2**(2), 91. DOI: 10.1016/S1566-2535(01)00026-4.
19. ***B. Oztaysi, S.C. Onar, K. Goztepe, C. Kahraman***
Soft Comput., 2015, 16 pp. (http://link.springer.com/article/10.1007/s00500-015-1853-8). DOI: 10.1007/s00500-015-1853-8.
20. ***K.T. Atanassov***
Fuzzy Sets and Systems, 1986, **20**(1), 87.
DOI: 10.1016/S0165-0114(86)80034-3.
21. ***J.M. Cadenas, M.C. Garrido, R. Martínez***
Expert Systems with Applications, 2013, **40**(16), 6241.
DOI: 10.1016/j.eswa.2013.05.051.
22. ***Y. Saeys, T. Abeel, Y. Van de Peer***
В *Proc. European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD'08),* Part II, Springer-Verlag Publ., 2008, pp. 313–325. DOI: 10.1007/978-3-540-87481-2_21.
23. ***L. Breiman***
Machine Learning, 2001, **45**(1), 5. DOI: 10.1023/A:1010933404324.
24. ***L. Breiman***
Machine Learning, 1996, **24**(1), 41. DOI: 10.1023/A:1018094028462.
25. ***J. Bergstra, Y. Bengio***
JMLR, 2012, **13**, 281.
26. ***C. Strobl, A.-L. Boulesteix, A. Zeileis, T. Hothorn***
BMC Bioinformatics, 2007, 21 pp. (http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-8-25).
DOI: 10.1186/1471-2105-8-25.
27. ***K. Archer***
Computational Statistics & Data Analysis, 2008, **52**(4), 2249.
DOI: 10.1016/j.csda.2007.08.015.
28. ***W. Webber, A. Moffat, J. Zobel***
ACM TOIS, 2010, **28**(4), Article No. 20.
DOI: 10.1145/1852102.1852106.
29. ***И.А. Тихомиров, И.В. Смирнов, И.В. Соченков, Д.А. Девяткин, А.О. Шелманов, Д.В. Зубарев, А.В. Швец, А.В. Лешкин, Р.Е. Суворов***
В *Труд. конф. Тринадцатая национальная конференция по искусственному интеллекту КИИ-2012 в 4 тт., (РФ, Белгород, 16–20 октября, 2012),* т. 4, Белгород, Изд. БГТУ, 2012, с. 100–108.

English

The Study of Scientific Projects Evaluation Criteria by Means of Machine Learning Methods Through the Examples of Russian Foundation for Basic Research Grant Competitions *

Dmitriy A. Devyatkin –
Institute for System Analysis FRC "Computer Science and Control" RAS
9, 60-letiya Octyabrya Ave., Moscow, 117312, Russia
e-mail: devyatkin@isa.ru

Roman E. Suvorov –
Institute for System Analysis FRC "Computer Science and Control" RAS
9, 60-letiya Octyabrya Ave., Moscow, 117312, Russia
e-mail: rsuvorov@isa.ru

Ilya A. Tikhomirov –
Institute for System Analysis FRC "Computer Science and Control" RAS
9, 60-letiya Octyabrya Ave., Moscow, 117312, Russia
e-mail: tih@isa.ru

Oleg G. Grigoriev –
Institute for System Analysis FRC "Computer Science and Control" RAS
9, 60-letiya Octyabrya Ave., Moscow, 117312, Russia
e-mail: ogrigoriev@gmail.com

Abstract

The paper considers the result of analysis of the main criteria for research projects evaluation by scientific foundations. A brief review of the experts' surveys analysis methods in making decision on projects funding is given. The novel approach to the determination of importance of scientific projects evaluation criteria is proposed; it is based on machine learning techniques. The approach allows scientist to apply the qualitative and quantitative criteria in case of large volume of data and does not require bringing the criteria to the numerical value. The authors experimentally determined the relevance of scientific criteria for projects estimation through the examples of Russian Foundation for Basic Research grant competitions. It has been established that the set of the most important criteria for projects evaluation remains virtually unchanged for all scientific fields besides the "Natural science methods in humanitarian sciences" area. The authors concluded that the suggested approach can be applied for the verification of experts' final estimate as well as for the check of the importance of the newly introduced criteria.

Keywords: funding criteria, importance of criteria, machine learning, random forest, Gini importance, decision making.

* *The work was financially supported by RFBR (project 14-29-05075).*

Images & Tables

Table 1. *The main criteria for the projects evaluation in some scientific funds*

Country	Organization/ Foundation	Budget (billion rub.)	Year	Main funding criteria
UK	EPSRC [2]	80,2	2009	1. Excellence. (A proposal that demonstrates excellence can be characterized by terms such as: novel, ambitious, timely, exciting, at the international forefront, adventurous, elegant, or transformative). 2. Importance. 3. Linked proposals. 4. Conformity with the state of the art. (Applicants are asked to set their proposal in context in terms of the current state of knowledge and other work under way in the field). 5. Management and planning. 6. The effectiveness of the proposed research. Applicants should prove that the funding going to be used effectively. (The proposal should clearly demonstrate the methodology, which the applicants intend to use to attain their stated objectives). 7. Activities aimed at promoting the engineering science, maths or physics.
Germany	DFG [3]	143,6	2013	1. Scientific quality of the project. 2. Applicants' qualifications. 3. Aims and work programme, planned allocation of funding. 4. The host institution and the research environment.
USA	NSF [4]	435.7	2012	1. What is the intellectual merit of the proposed activity? (How well does the activity advance discovery and understanding while promoting teaching, training, and learning?) 2. What are the broader (social) impacts of the proposed activity?
	NIH [5]	2026,9	2010	1. Significance. (Does the project address an important problem or a barrier to progress in the field?) 2. Investigator(s). (Are the PD(s)/PI(s), collaborators, and other researchers well suited to the project?) 3. Innovation. (Does the application challenge and seek to shift current research or clinical practice paradigms by utilizing novel theoretical concepts, approaches or methodologies, instrumentation, or interventions?) 4. Approach. (Are the overall strategy, methodology, and analyses well-reasoned and appropriate to accomplish the specific aims of the project?) 5. Environment. (Are the institutional support, equipment and other physical resources available to the investigators adequate for the project proposed?)
Europe Union (head office in France)	ESF [6]	3,6	2010	1. Relevance and expected impacts (driven by programme policy, strategy, mandates, etc.). 2. Scientific quality. 3. Social importance of the project. 4. Leader's qualification. The stage of development of management strategy. 5. The impact of this grant on related projects. 6. Cooperation with other research groups. 7. The expected synergistic effect.
Russian Federation	RFBR [7]	9.2	2014	1. Have the researchers successfully completed many projects before the application? 2. Is the project scientifically fundamental? 3. Are proposed methods reasonable? 4. Is the project feasible? 5. Are the problem and objectives clear? 6. Are the expected results important? 7. Is the project applied or theoretical?
	RSF [8]	11.4	2014	1. Is the topic of the application in the list of topics supported by the foundation? 2. What is the level of applicants' qualification? 3. Is the project feasible? 4. Are the expected results important? 5. Is the project's plan well done?
Finland	Suomen Akatemia [9]	24.6	2011	1. Scientific quality, innovativeness and novelty of the research plan. 2. Competence of the applicant / the research team. 3. Feasibility of the research plan. 4. Contacts with other research groups. 5. Project's significance for the promotion of the researchers' professional careers.

 DOI: 10.22204/2410-4639-2016-092-04-135-146 № 4 (92) октябрь–декабрь 2016 г.

Table 2. *The results of the experiments on the projects classification in the scientific-specific areas*

Scientific area	The number of requests	Accepted projects	Rejected projects	F_1, %	P, %	R, %
Math, computer science, mechanics	1859	581	1278	74.1	80.8	68.6
Physics, astronomy	2749	841	1908	67.0	80.0	57.6
Chemistry	2486	1734	752	69.5	82.7	60.2
Bio&Med sciences	4097	1265	2832	74.3	79.9	69.8
Earth sciences	2137	1479	658	66.1	78.3	57.7
Natural science methods in humanitarian sciences	1367	318	1049	69.8	71.3	84.9
Info-communication technologies and computer systems	1988	589	1399	76.1	82.9	70.8
Engineering sciences	2738	828	1910	69.4	79.5	62.6

Table 3. *The most important projects evaluation criteria in the scientific-specific areas*

Criterion	Rank of criterion in scientific-specific area							
	Math, computer science, mechanics	Physics, astronomy	Chemistry	Bio&Med sciences	Earth sciences	Natural science methods in humanitarian sciences	Info-communication technologies and computer systems	Engineering sciences
Have the researchers successfully completed many projects before the application?	1	1	2	1	3	7	1	1
Is the project scientifically fundamental?	2	4	6	6	6	1	3	4
Are the proposed methods reasonable?	3	3	1	2	1	4	5	2
Are the problem and objectives clear?	4	5	4	5	4	2	6	6
Are the expected results important?	5	2	3	4	5	5	4	3
Amount of publications of the project leader	6	7	7	9	8	-*	10	9
Academic degrees of the researchers	7	-	-	-	12	-	-	-
Is the project feasible?	8	6	5	3	2	3	2	5
Total amount of the researchers	9	12	8	-	-	12	7	-
The project leader's age	10	10	12	10	11	10	12	11
Academic ranks of the researchers	11	-	9	-	-	9	8	-
Amount of scientific papers of the researchers	12	11	11	12	9	-	-	10
Age of the researchers	-	8	-	11	7	11	11	8
Total amount of scientific papers of the researchers related to the project thematically	-	9	-	7	10	-	-	7
Requested funding	-	-	10	8	-	-	9	12
Research area	-	-	-	-	-	6	-	-
Official position of the researcher	-	-	-	-	-	8	-	-

Table 4. *The proximity of the most important criteria for the fundamental projects evaluation in accordance with* RBO_{12}*-criterion*

Research area	Math, computer science, mechanics	Physics, astronomy	Chemistry	Bio&Med sciences	Earth sciences	Natural science methods in humanitarian sciences	Info-communication technologies and computer systems	Engineering sciences
Math, computer science, mechanics	1.00	0.79	0.48	0.74	0.37	0.41	0.74	0.77
Physics, astronomy	0.79	1.00	0.54	0.79	0.37	0.24	0.73	0.87
Chemistry	0.48	0.54	1.00	0.60	0.80	0.24	0.42	0.64
Bio&Med sciences	0.74	0.79	0.60	1.00	0.55	0.29	0.78	0.90
Earth sciences	0.37	0.37	0.80	0.55	1.00	0.31	0.45	0.47
Natural science methods in humanitarian sciences	0.41	0.24	0.24	0.29	0.31	1.00	0.34	0.24
Info-communication technologies and computer systems	0.74	0.73	0.42	0.78	0.45	0.34	1.00	0.75
Engineering sciences	0.77	0.87	0.64	0.90	0.47	0.24	0.75	1.00

References

1. ***M. Benner, U. Sandström***
   Research Policy, 2000, **29**(2), 291. DOI: 10.1016/S0048-7333(99)00067-0.
2. *EPSRC – Engineering and Physical Sciences Research Council.* (https://www.epsrc.ac.uk/).
3. *DFG – Deutsche Forschungsgemeinschaft.* (http://www.dfg.de/en/).
4. *NSF – National Science Foundation.* (https://www.nsf.gov./).
5. *NIH – National Institutes of Health.* (https://www.nih.gov/).
6. *ESF – European Science Foundation.* (http://www.esf.org/).
7. *РФФИ – Российский фонд фундаментальных исследований.* (http://www.rfbr.ru/rffi/ru/).
8. RSF – Russian Science Foundation (in Russian). (http://grant.rscf.ru/).
9. *Academy of Finland.* (http://www.aka.fi/en/).
10. ***K. Olson***
    Field Methods, 2010, **22**(4), 295. DOI: 10.1177/1525822X10379795.
11. ***A. Geuna, B.R. Martin***
    Minerva, 2003, **41**(4), 277.
    DOI: 10.1023/B:MINE.0000005155.70870.bd.
12. ***J. Zhu, H. Wang, C. Ye, Q. Lang***
    Information Fusion, 2015, **24**(C), 93.
    DOI: 10.1016/j.inffus.2014.09.006.
13. ***O.I. Larichev***
    The Art and Science of Decision-Making, [Nauka i iskusstvo prinyatiya resheniy], RF, Moscow, Nauka Publ., 1979, 200 pp. (in Russian).
14. ***P. Weingart***
    Scientometrics, 2005, **62**(1), 117. DOI: 10.1007/s11192-005-0007-7.
15. ***A.B. Petrovskiy, G.V. Royzenzon, A.V. Balyshev, I.P. Tikhonov***
    IJ IMA, 2012, **1**(4), 349.
16. ***S. Wessely***
    The Lancet, 1998, **352**(9124), 301.
    DOI: 10.1016/S0140-6736(97)11129-1.
17. ***17. A.B. Petrovskiy***
    J. Information Technology and Computer Systems [Informatsionnye tekhnologii i kompyuternye sistemy], 2004, №2, 56 (in Russian).
18. ***A.-L. Jousselme, D. Grenier, E. Bossé***
    Information Fusion, 2001, **2**(2), 91. DOI: 10.1016/S1566-2535(01)00026-4.
19. ***B. Oztaysi, S.C. Onar, Kerim Goztepe, Cengiz Kahraman***
    Soft Comput., 2015, 16 pp. (http://link.springer.com/article/10.1007/s00500-015-1853-8). DOI: 10.1007/s00500-015-1853-8.
20. ***K.T. Atanassov***
    Fuzzy Sets and Systems, 1986, **20**(1), 87.
    DOI: 10.1016/S0165-0114(86)80034-3.
21. ***J.M. Cadenas, M.C. Garrido, R. Martínez***
    Expert Systems with Applications, 2013, **40**(16), 6241.
    DOI: 10.1016/j.eswa.2013.05.051.
22. ***Y. Saeys, T. Abeel, Y. Van de Peer***
    In *Proc. European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD'08),* Part II, Springer-Verlag Publ., 2008, pp. 313–325. DOI: 10.1007/978-3-540-87481-2_21.
23. ***L. Breiman***
    Machine Learning, 2001, **45**(1), 5. DOI: 10.1023/A:1010933404324.
24. ***L. Breiman***
    Machine Learning, 1996, **24**(1), 41. DOI: 10.1023/A:1018094028462.
25. ***J. Bergstra, Y. Bengio***
    JMLR, 2012, **13**, 281.
26. ***C. Strobl, A.-L. Boulesteix, A. Zeileis, T. Hothorn***
    BMC Bioinformatics, 2007, 21 pp. (http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-8-25).
    DOI: 10.1186/1471-2105-8-25.
27. ***K. Archer***
    Computational Statistics & Data Analysis, 2008, **52**(4), 2249.
    DOI: 10.1016/j.csda.2007.08.015.
28. ***W. Webber, A. Moffat, J. Zobel***
    ACM TOIS, 2010, **28**(4), Article No. 20.
    DOI: 10.1145/1852102.1852106.
29. ***I.A. Tikhomirov, I.V. Smirnov, I.V. Sochenkov, D.A. Devyatkin, A.O. Shelmanov, D.V. Zubarev, A.V. Shvets, A.V. Leshkin, R.E. Suvorov***
    In *Proc. 13th National Conference on Artificial Intelligence CAI-2012, 4-Vols Ed., [13-ya natsionalnaya konferentsiya po iskusstvennomu intellektu KII-2012], (RF, Belgorod, 16–20 October, 2012),* Vol. 4, Belgorod, BSTU Publ., 2012, pp. 100–108 (in Russian).

Подписано в печать 27.12.2016. Формат 60 x 90 $^{1}/_{8}$.
Печ. л. 18.5. Тираж 300 экз.

Оригинал-макет ЗАО «ИТЦ МОЛНЕТ»
123104, г. Москва, Малый Палашевский пер., д. 6
Тел./факс: (495) 927-01-98,
e-mail: info@molnet.ru
Печать ООО «ЛАЙФ»
105264, г. Москва,
7-я Парковая ул., д. 24, офис 100