На правах рукописи

Булатов Константин Булатович

Методы, модели и алгоритмы комбинирования и останова в системах распознавания в видеопотоке

Специальность 05.13.01 —
«Системный анализ, управление и обработка информации
(информационно-вычислительное обеспечение)»

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Москва — 2019

Работа выполнена в Федеральном государственном учреждении «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», отдел № 91.

Научный руководитель: канд. тех. наук
Арлазаров Владимир Викторович

Официальные оппоненты: **Соболевский Андрей Николаевич,**
доктор физико-математических наук,
ФГБУН Институт проблем передачи информации им. А.А. Харкевича Российской академии наук,
директор

Нейман-заде Мурад Искендер оглы,
кандидат физико-математических наук,
АО «МЦСТ»,
начальник отделения систем программирования

Ведущая организация: Федеральное государственное учреждение «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук» (ФГУ ФНЦ НИИСИ РАН)

Защита состоится ___ _____ 2019 г. в ___ часов на заседании диссертационного совета Д 002.073.04 на базе Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН) по адресу: 117312, г. Москва, проспект 60-летия Октября, 9.

С диссертацией можно ознакомиться в библиотеке библиотеке ФИЦ ИУ РАН по адресу: г. Москва, ул. Вавилова, д. 40 и на официальном сайте ФИЦ ИУ РАН: http://www.frccsc.ru.

Отзывы на автореферат в двух экземплярах, заверенные печатью учреждения, просьба направлять по адресу: 119333, г. Москва, ул. Вавилова, д. 44, кор. 2, ученому секретарю диссертационного совета Д 002.073.04.

Автореферат разослан ___ _____ 2019 г.
Телефон для справок: +7 (499) 135-51-64.

Ученый секретарь
диссертационного совета
Д 002.073.04,
д-р техн. наук, профессор В.Н. Крутько

Общая характеристика работы

Актуальность темы. Системы анализа и распознавания документов занимают значительное место в таких областях науки, как искусственный интеллект, теория принятие решений, и распознавание образов. Большой вклад в развитие данного научного направления внесли отечественные и зарубежные ученые М.А. Айзерман, В.Л. Арлазаров, Э.М. Браверман, Ю.В. Визильтер, И.Б. Гуревич, С.Ю. Желтов, Ю.И. Журавлев, А.Б. Мерков, А.Б. Петровский, В.А. Сойфер, Ян Лекун (Франция), Чэн-Линь Лю (КНР), Коити Кисэ (Япония), Джеффри Хинтон (Канада) и другие.

Использование смартфонов и планшетных компьютеров для решения задач оптимизации бизнес-процессов в корпоративных системах и процессов в системах государственного управления привели к новому витку развития систем компьютерного зрения, оперирующих на мобильных устройствах. Повышенный интерес к реализации корпоративного делопроизводства на основе мобильного документооборота, а также необходимость осуществления ввода документов в условиях с неконтролируемыми условиями съемки, повышают требования к системам распознавания, автоматического ввода и анализа документов с использованием мобильных устройств.

Изображения, полученные при помощи мобильных устройств, обладают рядом характерных особенностей и искажений, таких, как недостаточное разрешение, недостаточная либо неравномерная освещенность, смазывание, дефокусировка, блики на отражающей поверхности плоских объектов и другими. Подобные особенности входных изображений повышают требования к мобильным системам оптического распознавания и создают потребность в новых методах и алгоритмах, обладающих большей устойчивостью. Разработке методов распознавания образов, учитывающих особенности малоформатных цифровых камер, посвящены работы таких авторов, как Д.П. Николаев, О.А. Славин, Д.С. Ватолин, V. Lepetit, T. Geraud, R. Manmatha, D. Doermann, X. Bai, D. Karatzas, M. Iwamura и других. В то же время недостаточно изученными являются модели и методы использования видеопотока в качестве цифрового представления распознаваемого объекта, и методы повышения качества систем оптического распознавания путем использования множества гомогенных наблюдений распознаваемого объекта. Таким образом, дальнейшее исследование и развитие математических моделей и методов использования видеопотока в качестве цифрового представления объекта в контексте систем оптического распознавания является актуальным.

Основные результаты диссертации были получены в процессе выполнения работ по следующим научным грантам РФФИ:

– № 18-07-01387 – «Модели и методы построения систем оптического распознавания видеопотока с использованием обратных связей, функционирующим в в условиях ограниченных вычислительных ресурсов»;

– № 17-29-03370 – «Методы биометрической идентификации в реальном времени на мобильном устройстве по удостоверяющей фотографии»;

– № 17-29-03170 – «Исследование быстродействующих методов и алгоритмов обработки изображений и оптического распознавания для использования в мобильных устройствах с ограниченной вычислительной производительностью»;

– № 15-07-06520 – «Методы контроля подлинности документов и их фрагментов в гибридных системах обработки, передачи и хранения документов»;

– № 14-07-00730 – «Математическое моделирование шумовых помех при распознавании»;

– № 13-07-12172 – «Распознавание документов удостоверяющих личность с помощью веб камер и камер мобильных устройств».

Целью данной работы является разработка математических моделей, методов улучшения характеристик систем распознавания объектов в видеопотоке путем комбинирования результатов обработки множества входных наблюдений.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Провести анализ принципов построения современных систем распознавания документов;

2. Построить математическую модель системы распознавания объекта в видеопотоке, позволяющую исследовать качественные характеристики результата и время, необходимое для его получения;

3. Исследовать влияние характеристик входных данных на выбор оптимальной стратегии комбинирования результатов распознавания одиночных изображений;

4. Разработать алгоритм комбинирования результатов оптического распознавания строкового объекта и провести экспериментальный анализ его характеристик;

5. Разработать метод останова процесса распознавания объекта в видеопотоке в рамках построенной математической модели системы;

6. Разработать алгоритм останова процесса распознавания строкового объекта и провести экспериментальны анализ его характеристик;

7. Реализовать разработанные методы и алгоритмы для их внедрения в промышленные системы распознавания объектов в видеопотоке.

Методология и методы исследования основаны на системном анализе, математическом моделировании, математической статистике и теории принятия решений.

Основные положения, выносимые на защиту:

1. Построена математическая модель системы распознавания объекта в видеопотоке с модулем комбинирования результатов распознавания одиночных кадров и с модулем останова;

2. Показано преимущество правила максимальной оценки как стратегии комбинировании покадровых результатов классификации объекта в видеопоследовательностях, не содержащих ошибок локализации и сегментации объекта;

3. Разработан алгоритм комбинирования результатов распознавания строкового объекта, учитывающий альтернативные варианты классификации отдельных символов;

4. Разработан метод останова процесса распознавания объекта в видеопотоке на основе порогового отсечения оценки ожидаемого расстояния между текущим и следующим интегрированными результатами распознавания;

5. Разработан алгоритм останова процесса распознавания строкового объекта в видеопотоке с оценкой расстояния между текущим и следующим интегрированными результатами распознавания, вычисляемой по накопленным наблюдениям.

Научная новизна:

1. Предложена новая математическая модель системы распознавания объекта в видеопотоке, позволяющая проводить совместное исследование качественных характеристик результата распознавания и времени, необходимого для получения результата;

2. Выполнено оригинальное исследование влияния модели входных данных на выбор оптимальной стратегии комбинирования покадровых результатов, применительно к задаче классификации объекта в видеопотоке;

3. Разработан новый алгоритм комбинирования результатов распознавания строкового объекта, учитывающий альтернативные варианты классификации отдельных символов;

4. Предложен новый метод останова процесса распознавания произвольного объекта в видеопотоке, рассматривающий данный процесс как монотонную задачу останова и основывающийся на оценке ожидаемого расстояния между текущим и следующим интегрированными результатами;

5. Разработан новый алгоритм останова процесса распознавания строкового объекта в видеопотоке, основанный на оценке ожидаемого расстояния между текущим и следующим интегрированными результатами, вычисляемой по накопленным наблюдениям.

Практическая значимость. Разработанная в рамках диссертации модель системы распознавания объектов в видеопотоке, а также разработанные методы и алгоритмы комбинирования результатов распознавания строковых объектов и останова процесса распознавания были реализованы в виде программных компонентов и внедрены в программное обеспечение «Smart 3D OCR MRZ» и «Smart PassportReader» компании ООО «Смарт Энджинс РУС», а также «Smart IDReader» компании ООО «Смарт Энджинс Сервис». Данные продукты интегрированы в информационную ин-

фраструктуру ряда коммерческих организаций, а также в ряд информационных решений государственных структур Российской Федерации.

Достоверность полученных результатов подтверждается согласованностью разработанных алгоритмов, методов и математических моделей с экспериментальными результатами, представленными в работе, успешной апробацией результатов и внедрением в коммерческие системы распознавания документов.

Апробация работы. Основные результаты работы докладывались на следующих семинарах и конференциях:

1. 7th International Workshop on Camera Based Document Analysis and Recognition (CBDAR 2017), Киото, Япония, 2017;

2. 10th International Conference on Machine Vision (ICMV 2017), Вена, Австрия, 2017;

Личный вклад. Все результаты, изложенные в диссертации, принадлежат лично автору. В совместных работах автор принимал непосредственное участие в выборе направления и задач исследований, в построении математических моделей и обсуждении результатов экспериментальных исследований.

Публикации. Основные результаты по теме диссертации изложены в 13 публикациях, в том числе: 5 изданы в журналах, рекомендованных ВАК, 3 – в сборниках трудов конференций (входящих в международные базы цитирования Scopus и Web of Science), 2 патента на полезную модель и 3 свидетельства о государственной регистрации программы для ЭВМ.

Объем и структура работы. Диссертация состоит из введения, четырех глав и заключения. Полный объем диссертации составляет 107 страниц, включая 17 рисунков и 6 таблиц. Список литературы содержит 139 наименований.

Содержание работы

Во **введении** обосновывается актуальность диссертационной работы и ее научная новизна, формулируются основные цели и задачи диссертационного исследования, приводятся положения, выносимые на защиту, а также краткое содержание глав диссертации.

Первая глава посвящена анализу принципов построения современных систем распознавания документов. Рассматривается автоматический ввод документов как одна из основных задач, возникающих в рамках электронного и мобильного документооборота. Выделены основные этапы обработки изображений документов, характерных для систем автоматического ввода, такие, как поиск и локализация документов, сегментация изображений документов, распознавание одиночных объектов, пост-обработка, оценка достоверности распознавания.

Показано, что современные работы, связанные с автоматическим вводом и распознаванием документов на мобильных устройствах, рассматривают фотографию документа как его электронное представление и отмечают трудности, связанные с подготовкой образа документа к распознаванию и с самим распознаванием. В связи с этим целью диссертации является исследование видеопотока как цифрового представления распознаваемого объекта, возникающих в данном контексте задач использования множества входных наблюдений как способа повышения точности распознавания, а также изучение скорости получения результата распознавания объекта в видеопотоке как важного фактора эффективности системы.

Рассматривая время получения результата как характеристику системы распознавания объекта в видеопотоке в реальном времени, возникает важная задача, не присущая классическим системам распознавания на одиночных изображениях – задача останова процесса распознавания. Существует множество теоретически проработанных в литературе постановок этой задачи, включая задачу о разборчивой невесте, задачу о продаже дома, задачу о максимизации моментов и т.п. При этом задача останова процесса распознавания остается малоизученной и на данный момент не предложено универсальных методов, применяющих теорию оптимального останова к задаче распознавания объекта в видеопотоке.

Во **второй главе** предложена новая математическая модель системы оптического распознавания объекта в видеопотоке с модулем комбинирования результатов распознавания объекта на одиночных изображениях и с модулем останова. При использовании мобильных устройств возникает возможность рассматривать видеопоток как цифровое представление объекта, что позволяет решать задачи, недоступные в случае анализа одиночной фотографии. Внешние условия съемки могут привести к тому, что объект на одиночном кадре искажен. Примером такого искажения является блик, проявляющийся на глянцевой поверхности (см. рис. 1). Поскольку в видеопотоке геометрическое положение снимаемого объекта может меняться, блик также меняет свое положение, что позволяет получить информацию о скрываемом объекте на другом кадре видеопотока.

Предлагаемая модель системы распознавания объекта в видеопотоке представлена на рисунке 2. Видеопоток представляется как последовательность $I_t(x) \in \mathbb{I}$ захватываемых изображений объекта x в дискретные моменты времени t. Время, необходимое для получения обновленного результата распознавания после ввода очередного образа $I_t(x)$, в общем случае, является функцией от изображения и внутреннего состояния системы. Результат, учитывающий изображение $I_t(x)$, может быть доступен только в момент времени $T(t) \geq t$. Пусть в момент времени t_0 происходит захват изображения $I_{t_0}(x)$. Результат распознавания одиночного кадра $\hat{f}(I_{t_0}(x))$ становится доступным в момент времени $t_1 \geq t_0$ и регистрируется в модуле памяти системы. После этого происходит комбинирование результатов

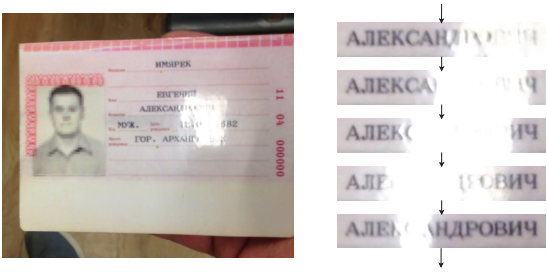


Рис. 1 — Фрагмент кадра с бликом на отражающей поверхности документа (слева) и извлеченные изображения текстового поля (справа).

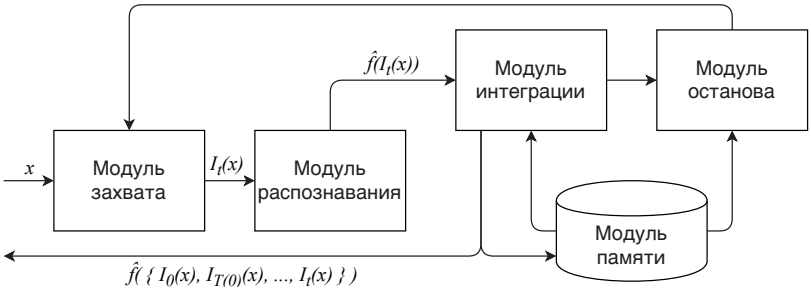
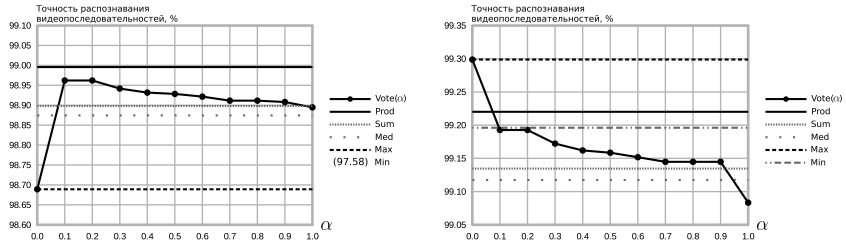


Рис. 2 — Схема системы распознавания в видеопотоке с остановом.

распознавания изображений объекта, накопленных на текущий момент, и в момент времени $t_2 = T(t_0) \geq t_1$ происходит вывод текущего интегрированного результата распознавания R_{t_2}, учитывающий информацию, которая содержится в изображениях $I_0(x), I_{T^1(0)}(x), I_{T^2(0)}(x), \ldots, I_{t_0}(x)$. Качество результата характеризуется близостью результата R_{t_2} к истинному значению $\nu(x)$ объекта x, согласно некоторой метрике. После вывода результата происходит захват очередного изображения $I_{t_2}(x)$ и процесс продолжается.

В рамках подобной системы возникают задачи, нетипичные для традиционных систем распознавания объектов: задача комбинирования (интеграции) результатов распознавания одного и того же объекта на разных изображениях в единый результат R_{t_2}, и задача останова процесса распознавания, т.е. принятия в момент времени t_2 решения о том, что процесс захвата следует прекратить и накопленный к текущему моменту результат принять за окончательный. В качестве функционала эффективности системы в момент останова $t = t_{\text{stop}}$ предлагается рассматривать линейную комбинацию $a \cdot \rho(R_{t_{\text{stop}}}, \nu(x)) + b \cdot W(t_{stop})$, где a, b – константы, $\rho(R_t, \nu(x))$ – расстояние от интегрированного результата R_t до истинного значения $\nu(x)$, характеризующая качество результата, а $W(t)$ – штрафная функция от времени. Частным случаем штрафной функции $W(t)$ является количество обработанных изображений $W(t) = \max\{i \mid T^i(0) \leq t\}$.

На построение модуля интеграции результатов распознавания отдельных изображений может влиять как модель результата распознавания, так и природа распознаваемого объекта. Важным вопросом в рамках этой задачи является вопрос о выборе стратегии, которая будет оптимальной при заданной модели входных данных. Для исследования влияния модели входных данных на выбор оптимальной стратегии комбинирования в рамках диссертационной работы был поставлен следующий эксперимент: были подготовлены наборы данных, содержащие видеопоследовательности отдельных печатных символов с искажениями, характерными для мобильной видеосъемки, и результаты их классификации при помощи сверточной нейронной сети, обученной на независимой обучающей выборке. В видеопоследовательностях, объединенных в группу A, содержались изображения с ошибками предварительной обработки (вызванные некорректной или недостаточно точной работой алгоритмов локализации документа, сегментации текстовых строк на отдельные символы и т.п.). В видеопоследовательностях группы B не содержалось ошибок предварительной обработки. На полученных наборах данных проведено сравнение базовых стратегий комбинирования классификаторов в рамках Байесовской модели результата классификации: правило произведения оценок, суммы оценок, минимума, максимума, медианы, а также обобщенное правило голосования: линейная комбинация частоты класса с коэффициентом α и его максимальной оценки с коэффициентом $(1 - \alpha)$.



а) Группа A б) Группа B

Рис. 3 — Сравнение точности распознавания видеопоследовательностей символов с использованием базовых стратегий комбинирования.

На рисунке 3 продемонстрирована значительная разница в оптимальном выборе стратегии комбинирования в зависимости от модели входных данных: на тестовых наборах группы A более высокую точность распознавания видеопоследовательностей обеспечивают правило произведения, голосование и правило суммы. При этом на тестовых наборах группы B, в которых не содержалось ошибок предварительной локализации и сегментации символов, более высокую точность распознавания обеспечивает правило максимума.

Следует отметить, что прямое применение рассмотренных правил комбинирования невозможно в случае, если модель результата распознавания объекта более сложна, чем простой результат классификации. В качестве примера такого объекта можно назвать текстовую строку (строковый объект), для которой классификация производится независимо для каждого символа.

Третья глава посвящена разработке алгоритма комбинирования (интеграции) результатов распознавания строкового объекта в видеопотоке в рамках модели результата, учитывающей альтернативные варианты классификации отдельных символов.

Результат распознавания одиночного символа рассматривался как отображение из множества классов C, объединенного с меткой пустого класса λ, в множество нормализованных оценок принадлежности. В качестве метрики на множестве $\hat{C}$ всевозможных результатов распознавания символа использовался вариант манхэттенской метрики с множеством значений, заключенным в отрезке $[0, 1]$.

Пусть задана функция r комбинирования результатов распознавания одиночных символов $r : \hat{C}^N \times (\mathbb{R}_0^+)^N \to \hat{C} \setminus \{\hat{\lambda}\}$, принимающая на вход N результатов распознавания одиночных объектов $a_1, a_2, \ldots, a_N$ таких, что $\exists i : a_i \neq \hat{\lambda}$, и набор ассоциированных с ними неотрицательных весов $w_1, w_2, \ldots, w_N$, отражающих значимость каждого из результатов. Потребуем от функции комбинирования r следующего свойства:

$$\begin{aligned} r(a_1, \ldots, a_N, w_1, \ldots, w_N) = \\ = r(r(a_1, \ldots, a_{N-1}, w_1, \ldots, w_{N-1}), a_N, w_1 + \ldots + w_{N-1}, w_N). \end{aligned} \quad (1)$$

В рамках предложенного алгоритма в качестве функции комбинирования одиночных символов использовалось взвешенное среднее:

$$r(a_1, \ldots, a_N, w_1, \ldots, w_N)(c) = \frac{1}{W_N} \sum_{i=1}^{N} a_i(c) \cdot w_i, \quad \forall c \in C \cup \{\lambda\}. \quad (2)$$

Обозначим как $\hat{\lambda}$ пустой результат классификации одиночного символа $\hat{\lambda} \stackrel{\text{def}}{=} \{(\lambda, 1), (c_1, 0), \ldots, (c_K, 0)\}$. Результатом X распознавания строкового объекта будем называть последовательность элементов, принадлежащих множеству $\hat{C} \setminus \{\hat{\lambda}\}$. В качестве метрики $\rho_{\mathbb{X}}$ на множестве $\mathbb{X}$ всевозможных результатов распознавания строчных объектов использовалось нормализованное обобщенное расстояния Левенштейна.

Пусть заданы N результатов распознавания строкового объекта $X_1, \ldots, X_N$, где $|X_i| = n_i > 0$:

$$X_1 = x_1^1 x_2^1 \ldots x_{n_1}^1, \quad X_2 = x_1^2 x_2^2 \ldots x_{n_2}^2, \quad \ldots, \quad X_N = x_1^N x_2^N \ldots x_{n_N}^N, \quad (3)$$

а также для каждого результата X_i задан его вес w_i. При расчете интегрированного результата распознавания строкового объекта будем порождать набор промежуточных результатов $R^{(i)}(X_1,\dots,X_i,w_1,\dots,w_i)$. На первом шаге алгоритма: $R^{(1)}(X_1,w_1) = X_1$. На каждом последующем i-м шаге алгоритма строится оптимальное выравнивание строк X_i и $R^{(i-1)}(X_1,\dots,X_{i-1},w_1,\dots,w_{i-1})$ при помощи схемы динамического программирования.

Пусть $d(l,m) \stackrel{\text{def}}{=} \rho_{\mathbb{X}}\left(X_{i1\dots l}, R^{(i-1)}(X_1,\dots,X_{i-1},w_1,\dots,w_{i-1})_{1\dots m}\right)$, а $P_p(l,m)$ – вспомогательные функции для $p \in \{1,2,3\}$. Расчет $d(l,m)$ и $P_p(l,m)$ производится согласно следующей процедуре:

$$
\begin{gathered}
d(0,0)=0, \quad d(l,0)=\sum_{k=1}^{l}\rho_{\hat{C}}(x_k^i,\hat{\lambda}), \quad d(0,m)=\sum_{k=1}^{m}\rho_{\hat{C}}(\hat{\lambda},r_k^{(i-1)}),\\
\begin{aligned}
P_1(l,m) &= \rho_{\hat{C}}(x_l^i,\hat{\lambda}) + d(l-1,m),\\
P_2(l,m) &= \rho_{\hat{C}}(\hat{\lambda},r_m^{(i-1)}) + d(l,m-1),\\
P_3(l,m) &= \rho_{\hat{C}}(x_l^i,r_m^{(i-1)}) + d(l-1,m-1),\\
d(l,m) &= \min\{P_1(l,m),P_2(l,m),P_3(l,m)\}.
\end{aligned}
\end{gathered}
\tag{4}
$$

Для расчета результата на i-м шаге $R^{(i)}(X_1,\dots,X_i,w_1,\dots,w_i)$ введем две вспомогательные функции $t_X : \{0,\dots,n_i+n_{R_{i-1}}\} \to \{1,\dots,n_i\}$ и $t_R : \{0,\dots,n_i+n_{R_{i-1}}\} \to \{1,\dots,n_{R_{i-1}}\}$, расчет которых производится по следующей рекуррентной процедуре:

$$
\begin{aligned}
&t_X(0)=n_i, \quad t_R(0)=n_{R_{i-1}},\\
&t_X(k+1)=\begin{cases} t_X(k), \text{ если } \begin{array}{l} P_2(t_X(k),t_R(k)) = d(t_X(k),t_R(k)) \wedge \\ \wedge\, P_1(t_X(k),t_R(k)) \neq d(t_X(k),t_R(k)) \end{array} \\ t_X(k)+1, \text{ в остальных случаях,} \end{cases}\\
&t_R(k+1)=\begin{cases} t_R(k), \text{ если } P_1(t_X(k),t_R(k)) = d(t_X(k),t_R(k)) \\ t_R(k)+1, \text{ в остальных случаях.} \end{cases}
\end{aligned}
\tag{5}
$$

Результат на i-м шаге рассчитывается следующим образом:

$$
\begin{aligned}
&n_{R_i} = \min\left\{k : t_X(k) = t_R(k) = 0\right\},\\
&R(X_1,\dots,X_i,w_1,\dots,w_i) = r_1 r_2 \dots r_{n_{R_i}},\\
&r_k = \begin{cases} r\left(r^{(i-1)}_{t_R(t(k))+1}, \hat{\lambda}, W_{i-1}, w_i\right), \text{ если } t_X(t(k)) = t_X(t(k)-1), \\ r\left(\hat{\lambda}, x^i_{t_X(t(k))+1}, W_{i-1}, w_i\right), \text{ если } t_R(t(k)) = t_R(t(k)-1), \\ r\left(r^{(i-1)}_{t_R(t(k))+1}, x^i_{t_X(t(k))+1}, W_{i-1}, w_i\right), \text{ в остальных случаях,} \end{cases}
\end{aligned}
\tag{6}
$$

где $W_i \stackrel{\text{def}}{=} \sum_{k=1}^{i} w_k$, вспомогательная функция $t(k) \stackrel{\text{def}}{=} n_{R_i} - k + 1$, а r – функция интеграции результатов распознавания одиночных объектов (2).

Трудоемкость вычисления функции r (2) и метрики на результатах распознавания одиночных символов составляет $O(K)$, где K – количество классов, на которое происходит классификация каждого символа. Поскольку верхняя оценка на длину результирующей строки R после выполнения i-й итерации алгоритма составляет $O\left(\sum_{j=1}^{i} |X_i|\right) \leq O\left(i \cdot \max_{j=1}^{i} |X_i|\right)$, трудоемкость каждой итерации алгоритма можно оценить как $O(M^2NK)$, где $M = \max_{i=1}^{N} |X_i|$, и общую трудоемкость предлагаемого алгоритма как $O(M^2N^2K)$. В форме псевдокода описанный алгоритм представлен в полном тексте диссертации как Алгоритм 1.

В рамках экспериментального исследования Алгоритм 1, работающий в рамках расширенной модели результата распознавания строкового объекта, был сравнен с алгоритмом ROVER, оперирующим только с максимальными альтернативами классификации одиночных объектов. Результаты сравнения представлены на рисунке 4. Оба алгоритма интеграции показывают увеличение точности результата распознавания с увеличением количества кадров, при этом Алгоритм 1 достигает меньшего значения ошибки, чем интеграция методом ROVER, вне зависимости от длины последовательности интегрируемых результатов.

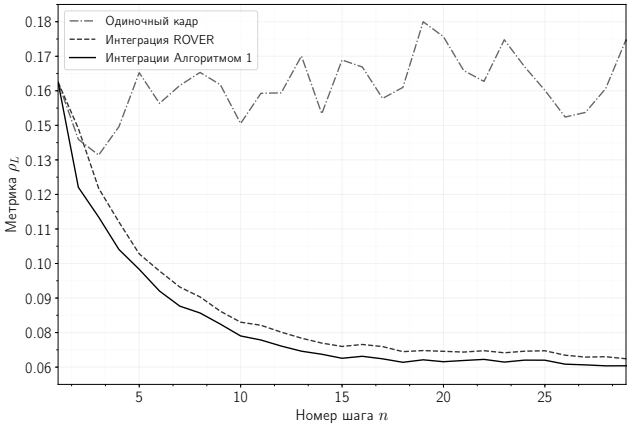


Рис. 4 — Результаты работы алгоритмов интеграции для текстовых полей пакета данных MIDV-500

Достигнутые средние значения расстояния между интегрированным результатом распознавания текстового поля и его истинным значениям для различных длин интегрируемого префикса видеопоследовательности представлены в таблице 1.

Таблица 1 — Достигнутое расстояние между интегрированным результатом распознавания и истинным значением без интеграции

Метод интеграции	Номер кадра					
	3	6	9	12	15	18
Без интеграции	0.136	0.154	0.160	0.157	0.168	0.159
Интеграция методом ROVER	0.125	0.096	0.083	0.075	0.070	0.069
Интеграция Алгоритмом 1	**0.115**	**0.089**	**0.078**	**0.071**	**0.066**	**0.065**

По форме графиков зависимости расстояния между интегрированным результатом и истинным значением от количества использованных кадров (см. рис. 4) можно судить о том, что интеграция обладает свойством убывающей доходности (в терминологии алгоритмов «anytime»). Это свойство является важным для решения задачи останова распознавания объекта в видеопоследовательности.

В **четвертой главе** предложен новый метод останова процесса распознавания объекта в видеопотоке на основе порогового отсечения оценки ожидаемого расстояния между текущим и следующим интегрированными результатами, и представлен новый алгоритм останова распознавания строкового объекта.

Пусть $\mathbb{X}$ обозначает множество всевозможных значений распознаваемого объекта, с заданной на нем метрикой ρ. В видеопотоке производится распознавание объекта с истинным значением $X^* \in \mathbb{X}$. Процесс распознавания предполагает, что наблюдается последовательность случайных результатов распознавания $\mathbf{X} = (X_1, X_2, \ldots)$, один результат за один шаг процесса, и каждое наблюдение $x_i \in \mathbb{X}$ является реализацией X_i. Будем считать, что X_1,$X_2, \ldots$ имеют одинаковое совместное распределение с X^*.

Пусть определена функция интеграции нескольких результатов распознавания: $R : \mathbb{X}^+ \to \mathbb{X}$, при помощи которой в любой момент n может быть получен интегрированный результат $R_n = R(x_1, \ldots, x_n)$. Процесс может быть остановлен в любое время $n > 0$ со следующей функцией штрафа:

$$L_n \stackrel{\text{def}}{=} \rho(R_n, X^*) + c \cdot n, \tag{7}$$

где c – стоимость наблюдения. Данная функция штрафа является уточнением функционала эффективности модели системы распознавания объекта в видеопотоке, описанной во второй главе.

Правило останова может быть представлено как случайная величина N (случайное время останова), распределение которой зависит от входных наблюдений. Задача состоит в выборе правила останова, доставляющего минимум функционалу ожидаемого убытка $V(N) = \mathrm{E}(L_N(X_1, \ldots, X_N))$.

Особым классом задач останова является класс *монотонных* задач, определяемый следующим образом: пусть A_n обозначает событие $\{L_n \leq \mathrm{E}_n(L_{n+1})\}$. Задача останова называется монотонной, если выполняется $A_0 \subset A_1 \subset A_2 \subset \ldots$. Для монотонных задач с конечным горизонтом оптимальным является «близорукое правило останова», останавливающее

процесс на шаге n, если текущее значение функции убытка не превосходит ожидаемого значения убытка при останове на шаге $n+1$.

Сформулируем следующее требование к функции интеграции R: ожидаемое расстояние между двумя соседними интегрированными результатами распознавания не возрастает со временем:

$$\mathrm{E}(\rho(R_n,R_{n+1})) \geq \mathrm{E}(\rho(R_{n+1},R_{n+2})) \quad \forall n > 0. \tag{8}$$

Пользуясь таким предположением о функции интеграции R можно показать, что задача останова с функцией убытка (7) становится монотонной начиная с некоторого шага. Действительно, обозначим через B_n событие $\{\mathrm{E}_n(\rho(R_n,R_{n+1}) \leq c)\}$ и рассмотрим задачу останова начиная с шага n, на котором событие B_n впервые произошло. События A_n, рассматриваемые в условии монотонности, принимают следующий вид:

$$\begin{aligned} A_n : \{\rho(R_n,X^*) + cn \leq \mathrm{E}_n(\rho(R_{n+1},X^*)) + cn + c\} = \\ = \{\rho(R_n,X^*) - \mathrm{E}_n(\rho(R_{n+1},X^*)) \leq c\}. \end{aligned} \tag{9}$$

При фиксированном X^*, на шаге n, пользуясь неравенством треугольника, можно получить соотношение между расстоянием от текущего результата распознавания до истинного значения, ожидаемым расстоянием до результата на следующем шаге и ожидаемым расстоянием от следующего результата до истинного значения:

$$\begin{aligned} \rho(R_n,X^*) \leq \mathrm{E}_n(\rho(R_n,R_{n+1})) + \mathrm{E}_n(\rho(R_{n+1},X^*)) \Rightarrow \\ \Rightarrow \rho(R_n,X^*) - \mathrm{E}_n(\rho(R_{n+1},X^*)) \leq \mathrm{E}_n(\rho(R_n,R_{n+1})). \end{aligned} \tag{10}$$

Если правая часть неравенства, полученного в (10) не превышает константы c, то и левая часть также не превышает c, и, следовательно, если происходит событие B_n, то и событие A_n (9) также должно произойти. Согласно предположению (8), если событие B_n произойдет, то и событие B_{n+1} также произойдет. Таким образом,

$$\forall n > 0 : \quad B_n \subset A_n, \quad B_n \subset B_{n+1}. \tag{11}$$

Из этого следует, что начиная с шага n, на котором событие B_n произошло впервые, события $A_n, A_{n+1}, A_{n+2} \ldots$ также произойдут, а значит задача останова может рассматриваться как монотонная начиная с этого шага, из чего следует оптимальность «близорукого» правила среди всех правил останова, достигающих шага n в случае, если задача имеет конечный горизонт.

Рассмотрим теперь правило останова, предписывающее останавливать процесс распознавания объекта в случае, если произошло событие B_n:

$$N_B = \min\{n > 0 : \mathrm{E}_n(\rho(R_n,R_{n+1})) \leq c\}. \tag{12}$$

Если правило N_B требует останова на шаге n, то и «близорукое» потребует останова на этом шаге, а поскольку проблема становится монотонной начиная с шага n, решение «близорукого» правила является оптимальным. Более того, если $\rho(R_n, X^*) - \mathrm{E}_n(\rho(R_{n+1}, X^*)) > c$, то правило N_B не останавливает процесс, также как и оптимальное правило. Следовательно, в случае если предположение (8) верно, правило N_B никогда не остановится раньше времени, и если правило требует останова, то решение об останове оптимально. На рисунке 5 графически показаны сходства и различия правил останова N_B и N^* при различных соотношениях между событиями A_n и B_n.

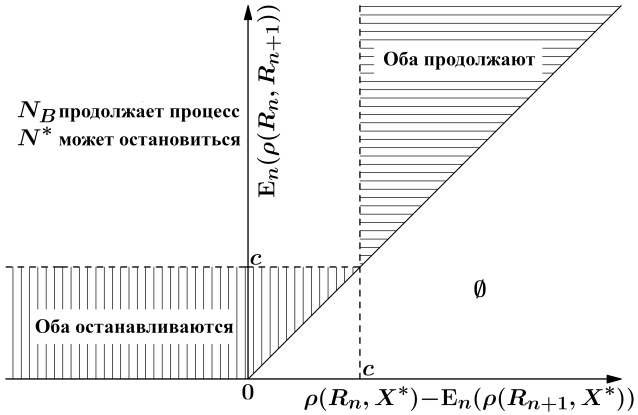


Рис. 5 — Разница поведений предлагаемого правила останова N_B и оптимального правила останова N^*

Тем самым, для решения задачи останова с функцией убытка (7) предлагается использовать следующий метод:

1. Оценить ожидаемое расстояние (в терминах метрики ρ) от текущего интегрированного результата распознавания объекта R_n (известного на шаге n) до неизвестного следующего результата R_{n+1};
2. Принимать решение об останове процесса на шаге n, производя пороговое отсечение расстояния, оцененного в пункте 1, таким образом аппроксимируя поведение правила N_B.

В общем случае выбор метода прогнозирования следующего интегрированного результата распознавания объекта (или оценки ожидаемого расстояния между ним и текущим интегрированным результатом) может зависеть от природы функции интеграции R и от других специфических характеристик задачи.

Построим на основе предложенного метода алгоритм останова процесса распознавания строкового объекта. Пусть задана функция интеграции результатов распознавания строкового объекта R (которая может быть

реализована при помощи метода ROVER или при помощи Алгоритма 1). В качестве метрики ρ на строковых объектах предлагается использовать нормализованное обобщенное расстояние Левенштейна. Для того, чтобы аппроксимировать поведение правила останова N_B, на n-м шаге процесса необходимо вычислять оценку ожидаемого расстояния между соседними интегрированными результатами распознавания $\Delta_n \stackrel{\text{def}}{=} \mathrm{E}(\rho(R_n, R_{n+1}))$, имея доступ к наблюдениям $X_1 = x_1, \ldots, X_n = x_n$. Для вычисления оценки предлагается провести моделирование следующего интегрированного результата исходя из предположения, что новое наблюдение будет близко к уже полученным на предыдущих шагах наблюдениям:

$$\hat{\Delta}_n \stackrel{\text{def}}{=} \frac{1}{n+1}\left(\delta + \sum_{i=1}^{n} \rho(R_n, R(x_1,x_2,\ldots,x_n,x_i))\right), \tag{13}$$

где δ – внешний настраиваемый параметр.

При использовании модели результата распознавания строкового объекта с альтернативными вариантами классификации одиночных объектов, рассмотренной в третьей главе, верхняя оценка длин интегрированных результатов R_N и R_{N+1} составляет $O(MN)$, где $M = \max_{i=1}^{N} |X_i|$. Поскольку трудоемкость прямого вычисления нормализованного обобщенного расстояния Левенштейна составляет $O(|X| \cdot |Y| \cdot K)$, где K – количество классов, на которое происходит классификация каждого одиночного объекта, трудоемкость алгоритма не превышает $O(M^2N^3K)$. Следует отметить, что трудоемкость алгоритма может быть снижена как путем использования упрощенных моделей результата распознавания строкового объекта, так и при помощи эвристических алгоритмов приближенного вычисления обобщенного расстояния Левенштейна. В форме псевдокода алгоритм представлен в полном тексте диссертации как Алгоритм 2.

Для оценки эффективности правила останова был построен профиль эффективности, графически показывающий зависимость среднего количества обработанных наблюдений и соответствующего среднего расстояния от полученного интегрированного результата в момент останова до истинного значения, при изменении стоимости наблюдения c. В качестве контрольного правила использовалось простое правило подсчета N_K, которое требует останавливать процесс распознавания на фиксированном шаге K. Дополнительно исследовались два варианта ранее опубликованного правила останова, основанного на пороговом отсечении размера наибольшего кластера идентичных результатов, накопленных к моменту n. Таким образом, построено два контрольных правила останова: N_{CX}, производящий пороговое отсечение размера наибольшего кластера идентичных результатов покадрового распознавания $x_1, \ldots, x_n$, и N_{CR}, аналогично рассматривающий интегрированные результаты распознавания $R_1, \ldots, R_n$.

Рисунок 6 иллюстрирует эффективность правил останова для текстовых полей пакета данных MIDV-500, распознаваемых при помощи библио-

теки Tesseract (версии 3.05.01 и 4.0.0). Интеграция результатов распознавания текстовых строк производилась методом ROVER. Более низкое положение кривой отражает большую эффективность правила останова. Можно отметить, что с среднем предлагаемое правило останова N_B (12) обладает большей эффективностью, чем другие исследованные методы. Кроме того, метод останова N_B (12) обладает высокой эффективностью без каких-либо модификаций для двух различных версий библиотеки Tesseract, использующих различные поколения алгоритмов распознавания текстовой строки.

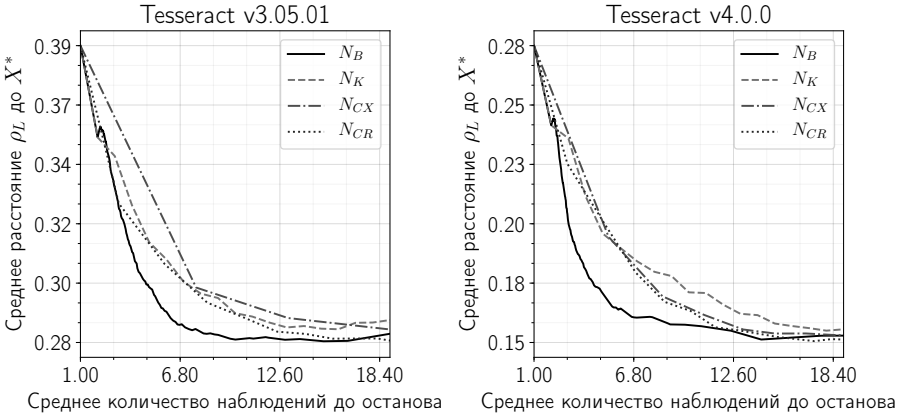


Рис. 6 — График зависимости среднего расстояния между полученным результатом в момент останова и истинным значением от среднего количества обработанных кадров до останова, при изменяющейся стоимости наблюдения c. Значение параметра $\delta = 0.2$

В таблице 2 показано среднее расстояние от интегрированного результата до правильного ответа в момент останова, которое может быть достигнуто при помощи исследованных правил останова, при распознавании текстовых полей при помощи библиотеки Tesseract v4.0.0. Колонки таблицы 2 отражают целевые интервалы для значений среднего количество использованных наблюдений (т.е. среднего количества обработанных кадров), строки таблицы соответствуют правилам останова, и каждая ячейка содержит результат замера с наименьшим средним количеством наблюдений, попадающим в данный интервал. Ячейки таблицы не содержат данных (помечены символом $\varnothing$) в случае, если соответствующее правило останова не способно достичь среднего количества обработанных кадров в соответствующем интервале для рассматриваемого набора входных данных. Можно сделать вывод, что во всех целевых интервалах правило останова N_B (12), разработанное в рамках диссертации, показывает лучший результат среди исследуемых альтернатив.

Таблица 2 — Достигнутые значения среднего расстояния от интегрированного результата до идеального значения в момент останова

Правило останова	Измеряемый параметр	Целевой интервал среднего количества наблюдений $E(N)$					
		3 ± 0.5	4 ± 0.5	5 ± 0.5	6 ± 0.5	7 ± 0.5	8 ± 0.5
N_{CX}	$\mathrm{E}(N)$ $\mathrm{E}(\rho_L(R_N, X^*))$	∅	∅	5.332 0.195	∅	∅	8.471 0.170
N_{CR}	$\mathrm{E}(N)$ $\mathrm{E}(\rho_L(R_N, X^*))$	2.936 0.227	∅	5.099 0.201	∅	6.920 0.180	∅
N_K	$\mathrm{E}(N)$ $\mathrm{E}(\rho_L(R_N, X^*))$	3.000 0.237	4.000 0.213	5.000 0.197	6.000 0.191	7.000 0.185	8.000 0.180
$\boldsymbol{N_B}$	$\mathrm{E}(N)$ $\mathrm{E}(\rho_L(R_N, X^*))$	**2.580** **0.224**	**3.551** **0.188**	**4.571** **0.174**	**5.539** **0.165**	**6.683** **0.161**	**7.742** **0.161**

В **заключении** приведены основные результаты работы, которые заключаются в следующем:

1. Построена математическая модель системы распознавания объекта в видеопотоке с модулем комбинирования покадровых результатов распознавания и с модулем останова. В качестве функционала эффективности системы была рассмотрена линейная комбинация расстояния от интегрированного результата распознавания до истинного значения объекта и штрафной функции от времени от момента начала процесса съемки до останова. Данная модель позволяет рассматривать систему распознавания объекта в видеопотоке как итерационный вычислительный процесс, который способен выдать в любое время наилучшее на данный момент решение, и прекратить захват новых изображений согласно заданному правилу останова.

2. Выполнено оригинальное исследование влияния модели входных данных на выбор оптимальной стратегии комбинирования покадровых результатов классификации в рамках задачи распознавания одиночного символа в видеопотоке. Показано, что если в последовательности обрабатываемых изображениях одиночного изображения отсутствуют ошибки предварительной обработки (такие, как ошибки локализации и сегментации символов), более высокую точность финального результата обеспечивает правило максимальной оценки. Для видеопоследовательностей, в которых встречаются ошибки локализации и сегментации символов, более высокую точность финального результата обеспечивают правила произведения оценок, правило голосования и правило суммы оценок.

3. Разработан новый алгоритм комбинирования результатов распознавания строкового объекта, учитывающий альтернативные варианты классификации отдельных символов (компонентов строкового объекта). Экспериментально показано, что предложенный алгоритм способен обеспечить более высокую точность интегрированного результата по сравнению с методом интеграции результатов распознавания как строк над множеством классов значений компонентов, применительно к задаче распознавания текстовой строки в видеопотоке.

4. Была рассмотрена задача останова процесса распознавания объекта в видеопотоке, что является важной и новой задачей, особенно актуальной при разработке систем оптического распознавания, предназначенных для работы на мобильных устройствах. Разработан новый метод останова процесса распознавания объекта в видеопотоке на основе порогового отсечения оценки ожидаемого расстояния между текущим и следующим интегрированными результатами распознавания. Метод разработан исходя из предположения о том, что задача останова процесса распознавания с интегрированием покадровых результатов становится монотонной начиная с некоторого шага. Справедливость данного предположения показано экспериментально. На основе разработанного метода предложен новый алгоритм останова процесса распознавания строкового объекта в видеопотоке, в котором оценка расстояния между текущим и следующим интегрированными результатами вычисляется путем моделирования следующего интегрированного результата с использованием уже накопленных наблюдений. Было продемонстрировано, что в задаче распознавания текстовых строк предложенное правило останова является более эффективным, чем пороговое отсечение количества обработанных кадров или пороговое отсечение размеров максимального кластера идентичных результатов.

5. Результаты работы в качестве программных компонентов систем распознавания документов в видеопотоке были внедрены в программное обеспечение «Smart 3D OCR MRZ» и «Smart PassportReader» компании ООО «Смарт Энджинс РУС», а также «Smart IDReader» компании ООО «Смарт Энджинс Сервис». Данные продукты интегрированы в информационную инфраструктуру ряда коммерческих организаций, а также в ряд информационных решений государственных структур Российской Федерации.

Публикации автора по теме диссертации

В изданиях из списка ВАК РФ

1. Модель системы распознавания в видеопотоке мобильного устройства / В.В. Арлазаров, К.Б. Булатов, А.В. Усков // Труды ИСА РАН, – 2018, – Т. 68, Спецвыпуск № S1, С. 73–82.
2. Выбор оптимальной стратегии комбинирования покадровых результатов распознавания символа в видеопотоке / К.Б. Булатов // Информационные технологии и вычислительные системы, – 2017, – № 3, – С. 45–55.
3. 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 // Вестник РФФИ, – 2016, – № 4, – С. 109–115.

4. 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 // Вестник РФФИ, – 2016, – № 4, – С. 97–108.
5. Проблемы распознавания машиночитаемых зон с использованием малоформатных цифровых камер мобильных устройств / К.Б. Булатов, Д.А. Ильин, Д.В. Полевой, Ю.С. Чернышова // Труды ИСА РАН, – 2015, – Т. 65, – № 3, – С. 85–93.

В сборниках трудов конференций

6. Optimal frame-by-frame result combination strategy for OCR in video stream / K. Bulatov, A. Lynchenko, V. Krivtsov // ICMV 2017. – International Society for Optics, Photonics, – 106961Z, – 2018 (Web of Science, Scopus).
7. Method of determining the necessary number of observations for video stream documents / V. Arlazarov, K. Bulatov, T. Manzhikov, O. Slavin, I. Janiszewski // ICMV 2017. – International Society for Optics, Photonics, – 106961X, – 2018 (Web of Science, Scopus).
8. Smart IDReader: Document recognition in video stream / K. Bulatov, V. Arlazarov, T. Chernov, O. Slavin, D. Nikolaev // ICDAR 2017, – 2017, – V. 6, – P. 39–44 (Web of Science, Scopus).

Патенты и свидетельства о регистрации программ для ЭВМ

9. Система распознавания документов в видеопоследовательности: патент РФ на полезную модель № 159733 / В.Л. Арлазаров, К.Б. Булатов, Д.П. Николаев, Д.В. Полевой, О.А. Славин, опубл. 20.02.2016 по заявке № 2015145155 от 21.10.2015.
10. Система распознавания изображений символов на основе обучающей выборки: патент РФ на полезную модель № 161580 / В.Л. Арлазаров, К.Б. Булатов, Д.А. Ильин, Д.П. Николаев, Т.С. Чернов, А.В. Шешкус, опубл. 27.04.2016 по заявке № 2015148233 от 10.11.2015.
11. Программа для распознавания идентификационных карт личности «Smart IDReader»: свидетельство о государственной регистрации программ для ЭВМ № 2016616961 / В.В. Арлазаров, Д.П. Николаев, С.А. Усилин, К.Б. Булатов, Т.С. Чернов, Д.Г. Слугин, Д.А. Ильин, П.В. Безматерных, А.А. Муковозов, Е.Е. Лимонова, опубл. 22.06.2016 по заявке № 2016612014 от 01.03.2016.
12. Библиотека для распознавания машиночитаемых строк в видеопотоке «Smart 3D OCR MRZ»: свидетельство о государственной регистрации программы для ЭВМ № 2015615712 / В.В. Арлазаров, К.Б. Булатов, А.Г. Волков, Д.А. Ильин, А.В. Куроптев, А.Е. Марченко, Д.П. Никола-

ев, Д.В. Полевой, Т.С. Чернов, Ю.С. Чернышова, опубл. 22.05.2015 по заявке № 2015612888 от 10.04.2015.

13. Библиотека для распознавания в видеопотоке паспорта гражданина Российской Федерации «Smart PassportReader»: свидетельство о государственной регистрации программы для ЭВМ № 2015616071 / В.В. Арлазаров, К.Б. Булатов, Д.А. Ильин, А.В. Куроптев, Д.П. Николаев, Д.В. Полевой, С.А. Усилин, И.А. Фараджев, Т.С. Чернов, опубл. 29.05.2015 по заявке № 2015612880 от 10.04.2015.

Булатов Константин Булатович

Методы, модели и алгоритмы комбинирования и останова в системах распознавания в видеопотоке

Автореф. дис. на соискание ученой степени канд. тех. наук

Подписано в печать ___.___._____. Заказ № _____
Формат 60×90/16. Усл. печ. л. 1. Тираж 100 экз.
Типография _____