

На правах рукописи

Бабенко Артем Валерьевич

**ЭФФЕКТИВНЫЕ АЛГОРИТМЫ ПОИСКА
ПО БОЛЬШИМ КОЛЛЕКЦИЯМ
ИЗОБРАЖЕНИЙ**

Специальность 05.13.11 —
«Математическое и программное обеспечение вычислительных
машин, комплексов и компьютерных сетей»

Автореферат
диссертации на соискание учёной степени
кандидата физико-математических наук

Москва — 2017

Работа выполнена в федеральном государственном автономном образовательном учреждении высшего профессионального образования «Московский физико-технический институт (государственный университет)»

Научный руководитель: кандидат физико-математических наук,
профессор Сколковского института науки и технологий

Лемпицкий Виктор Сергеевич

Официальные оппоненты: **Визильтер Юрий Валентинович**,
доктор физико-математических наук,
ФГУП «ГосНИИАС»,
старший научный сотрудник

Браславский Павел Исаакович,

кандидат технических наук,
институт естественных наук и математики Уральского
Федерального Университета

Ведущая организация: Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный университет имени М.В. Ломоносова»

Защита состоится « » _____ 2017 г. в ____ часов на заседании диссертационного совета Д 002.024.01 на базе Федерального государственного учреждения «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук» по адресу: 125047, Москва, Миусская пл., д. 4.

С диссертацией можно ознакомиться в библиотеке ИПМ им. М.В.Келдыша РАН, <http://keldysh.ru>.

Автореферат разослан « » _____ 2017 г.

Ученый секретарь диссертационного совета

кандидат физико-математических наук

А. Е. Бондарев

Общая характеристика работы

Актуальность темы. Благодаря повсеместному распространению цифровых фотоаппаратов и видеокамер, а также популярности различных социальных сетей, в Интернете ежедневно появляются миллионы новых фотографий. В связи с этим, современным поисковым системам необходимо работать с коллекциями из миллиардов изображений. Для эффективной индексации и поиска по коллекциям такого объема приходится создавать специальные алгоритмы, позволяющие за несколько миллисекунд отвечать на пользовательские запросы.

Эффективные методы поиска по коллекциям изображений необходимы не только в работе поисковых систем, но и в более узких областях, таких как распознавание лиц, диагностика заболеваний на основе медицинских изображений, приложения для туризма и другие. Ключевым элементом каждого из этих приложений является поиск похожих, при котором для запроса необходимо найти визуально близкие изображения из большой коллекции.

Типичная схема работы системы визуального поиска представлена на рисунке 1. Поисковая система получает на вход изображение-запрос, затем представляет его в виде одного или нескольких векторов фиксированной длины. Эти векторы принято называть *дескрипторами* изображения. Все изображения из поисковой коллекции также представлены в ней в виде своих дескрипторов. Далее система находит изображения, дескрипторы которых оказались ближайшими к дескриптору запроса в смысле евклидовой метрики, и возвращает их пользователю в качестве результатов поиска. Описанный протокол работы визуального поиска сформировался в начале 90-ых годов и действует в настоящее время в современных сервисах Яндекс и Google.

Существует три основных направления исследований, актуальных для задач поиска по большим коллекциям изображений.

1. *Получение дескрипторов.* В первую очередь необходимо предложить алгоритм построения дескрипторов. Важно, чтобы дескрипторы изображений, которые содержат семантически похожие объекты или сцены, были близки в смысле евклидовой метрики многомерного евклидового пространства. Самые первые системы визуального поиска использовали простейшие дескрипторы: гистограммы цветов, описания текстур, информацию о границах и формах объектов

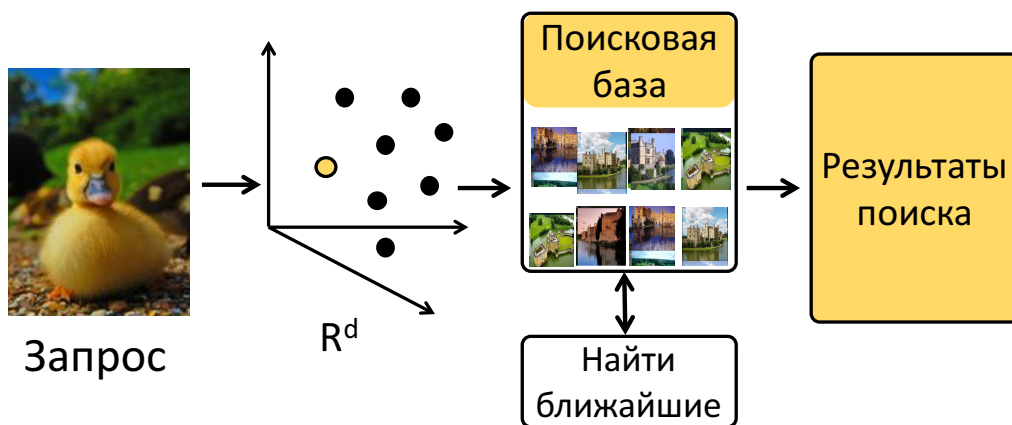


Рис. 1 — Типичная схема поиска по изображениям в современных поисковых системах.

на изображении. Такие тривиальные дескрипторы не могли обеспечить инвариантности к изменениям освещенности, поворотам, изменениям масштаба и угла зрения, необходимых для качественного поиска. В 1999 году были предложены СИФТ дескрипторы (SIFT, scale-invariant feature transform), основанные на гистограммах градиентов интенсивности и в некоторой степени обладавшие необходимыми типами инвариантности. Благодаря своей простоте и неплохому качеству работы СИФТы вплоть до недавнего времени были основным инструментом получения дескрипторов для визуального поиска. Для нескольких регионов изображения вычислялись локальные СИФТ векторы, и полученное множество являлось глобальным дескриптором изображения. В случае больших коллекций хранение всего множества локальных векторов для каждого изображения невозможно, так как их суммарный размер может превышать размер оперативной памяти поисковых серверов. Для решения этой проблемы было разработано несколько методов агрегации множества локальных векторов в один глобальный вектор-дескриптор. К сожалению, существующие глобальные дескрипторы имеют довольно высокую размерность порядка десятков тысяч, что делает невозможным их использование для больших по объему коллекций изображений.

2. *Сжатие дескрипторов.* Если коллекция содержит миллиарды изображений, то, даже в случае глобальных дескрипторов, их суммарный объем может намного превышать объем доступной оперативной памяти. Поэтому приходится сжимать имеющиеся дескрипторы и представлять каждый из них в виде компактного кода размером до нескольких байт. Такие компактные коды должны предоставлять возможность эффективно вычислять евклидово расстояние между несжатым дескриптором пользовательского запроса и сжатым дескриптором изображения из поисковой коллекции. Исторически первым подходом для решения этой задачи был подход семантического хэширования, предложенный в 2007 году. В этом подходе действительнoзначный вектор-дескриптор отображался в бинарную строку, причем близкие в исходном пространстве векторы отображались в строки, расстояние Хэмминга между которыми мало. Практическое преимущество этого подхода заключается в том, что вычисления на бинарных строках, как правило, гораздо быстрее операций с плавающей точкой. В 2011 году был предложен альтернативный подход к задаче сжатия, основанный на идее *мультиквантизации*. Этот подход эффективно аппроксимирует каждый вектор конкатенацией небольшого числа базовых векторов-слов из фиксированных множеств, а номера слов являются кодом вектора. Существенным недостатком подхода мультиквантизации является неявное предположение об отсутствии корреляций между различными размерностями дескрипторов, которое часто не выполняется для реальных данных.
3. *Поиск ближайших.* Наконец, для данного дескриптора запроса необходимо эффективно находить ближайших соседей среди миллиардов дескрипторов изображений поисковой коллекции. На таких объемах полный перебор невозможен, поэтому приходится использовать более быстрые приближенные алгоритмы. Приближенный поиск ближайших соседей — очень популярная задача в современном анализе данных. Большинство существующих методов избегают полного перебора, разбивая поисковое пространство на большое число непересекающихся регионов. После получения запроса ал-

горитм проверяет только те регионы, которые являются наиболее перспективными для этого запроса. Примерами самых известных подходов для решения этой задачи являются разделяющие деревья поиска, иерархический метод k-средних, локально-чувствительное хэширование. Все перечисленные методы требуют существенных затрат по памяти, поэтому неприменимы к нахождению ближайших соседей среди миллиардов векторов. Первый подход, применимый к задачам такого масштаба, был предложен в 2011 году и демонстрирует приемлемую точность поиска со временем работы порядка ста миллисекунд на запрос. Для многих приложений в реальном времени такой скорости недостаточно, поэтому построение более эффективных методов является важной исследовательской задачей.

Целью данной работы является разработка алгоритмов для эффективной системы визуального поиска, позволяющей осуществлять поиск по коллекциям, содержащих миллиарды изображений за несколько миллисекунд.

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

1. Исследовать существующие методы построения дескрипторов изображений и разработать метод, формирующий дескрипторы, обеспечивающие приемлемую точность поиска, размерность которых не превышает нескольких сотен.
2. Исследовать существующие методы сжатия векторов высокой размерности, оценить границы их применимости, разработать новый метод сжатия с меньшими потерями без увеличения бюджета потребляемой памяти.
3. Разработать новый алгоритм поиска ближайших векторов, способный обрабатывать запросы к коллекциям из миллиарда векторов за несколько миллисекунд с бюджетом по памяти несколько байт на вектор.

Научная новизна:

1. Предложены новые методы построения глобальных дескрипторов изображений для поиска, использующие, в отличие от аналогов,

глубокие сверточные нейросети, что позволяет превзойти существующие подходы к построению дескрипторов по точности поиска.

2. Предложены новые методы сжатия векторов высокой размерности, использующие квантизацию с неортогональными словарями и превосходящие существующие методы сжатия по точности кодирования.
3. Разработаны новые структуры данных, способные искать по коллекциям из миллиарда векторов за несколько миллисекунд с высокой точностью.

Теоретическая и практическая значимость. Теоретическая значимость работы заключается в предложенных структурах данных для эффективного нахождения ближайших соседей. Эти структуры могут быть использованы в качестве составляющих более сложных систем, например, для непараметрической классификации или непараметрического восстановления плотности. Также для этих структур получены оценки алгоритмической сложности основных операций. Практическая значимость работы заключается в разработке целостной системы визуального поиска, включающей все основные компоненты: построение дескрипторов, сжатие, эффективное нахождение соседей. Предложенные подходы используются в поиске по изображениям компании ООО "Яндекс".

Основные положения, выносимые на защиту:

1. Разработано два новых метода построения дескрипторов изображений на основе глубоких сверточных нейросетей. Во-первых, исследовано качество дескрипторов с полносвязных слоев предобученной нейросети и предложен способ повышения качества путем адаптации используемой нейросети к конкретной поисковой задаче. Также показано, что качество предложенных дескрипторов практически не снижается при сжатии их методом главных компонент. Во-вторых, предложен новый метод извлечения дескрипторов со сверточных слоев нейросети, преимущество которого заключается в том, что дескрипторы могут быть вычислены для изображения любого размера и формы. На большом количестве тестовых коллекций показано, что предложенные методы формируют более компактные

и семантически значимые дескрипторы по сравнению с существующими подходами.

2. Предложена новая модель аппроксимации векторов с помощью суммы слов, в отличие от конкатенации в модели мультиквантизации. На основе этой модели разработано два новых алгоритма сжатия векторов высокой размерности. Первый алгоритм — Аддитивная квантизация — не накладывает ограничений на значения слов и в теории приводит к наивысшему качеству сжатия, но требует решения сложной оптимизационной задачи, что делает кодирование неэффективным при больших длинах кода. Вторым алгоритмом — Древесная квантизация — накладывает определенные ограничения на слова, что снижает количество настраиваемых параметров, но обладает высокой скоростью кодирования. Доказана теорема о том, что ошибка реконструкции, достигаемая предложенными методами, не превосходит ошибки, достигаемой существующими подходами мультиквантизации. Экспериментально показано, что разработанные методы достигают лучшего качества сжатия по сравнению с передовыми алгоритмами для данных различной природы.
3. Разработаны ортогональный и неортогональный инвертированные мультииндексы — две структуры данных для эффективного поиска ближайших соседей, способные искать по миллиардам высокоразмерных векторов за несколько миллисекунд. Мультииндексы разбивают поисковое пространство на большое число регионов определенной структуры, позволяющей эффективно находить наиболее перспективные регионы для заданного запроса. Комбинация эффективности и большого числа регионов в разбиении позволяет обрабатывать запросы за несколько миллисекунд при высоком уровне точности поиска, что показано экспериментально. Также доказан теоретический результат об эффективности основных операций в предложенных структурах данных.

Достоверность полученных результатов обеспечивается доказанными теоретическими результатами, а также детально описанными протоколами экспериментов, которые могут быть воспроизведены. Программная реали-

зация предложенных алгоритмов и все данные для экспериментов доступны в сети Интернет.

Апробация работы. Основные результаты работы докладывались на:

1. 54-я научная конференция МФТИ, Долгопрудный, 2011, тема доклада: "Эффективный алгоритм поиска ближайших соседей при больших объемах поисковой базы"
2. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence RI, USA, 2012, тема доклада: "The inverted multi-index"
3. 56-я научная конференция МФТИ, Долгопрудный, 2013, тема доклада: "Компактное кодирование векторов большой размерности для быстрого поиска ближайших соседей"
4. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, USA, 2014, тема доклада: "Additive quantization for extreme vector compression"
5. IEEE European Conference on Computer Vision (ECCV), Zurich, Switzerland, 2014, тема доклада: "Neural codes for image retrieval"
6. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA, 2015, тема доклада: "Tree quantization for large-scale similarity search and classification"
7. Научный семинар Школы анализа данных Яндекса, Москва, 2015, тема доклада: "Древесная квантизация для компактного хранения векторов".
8. IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, тема доклада: "Aggregating local deep features for image retrieval"
9. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016, тема доклада: "Efficient Indexing of Billion-Scale Datasets of Deep Descriptors"
10. Научно-исследовательский семинар им. М. Р. Шура-Бура, ИПМ им. М.В.Келдыша, Москва, 2016, тема доклада: "Инвертированный мультииндекс"

Личный вклад диссертанта состоит в исследовании существующих методов, разработке предложенных алгоритмов и структур данных, доказательстве теоретических результатов, программной реализации предложенных подходов.

Структура и объем работы. Диссертация состоит из введения, трех глав, заключения и списка литературы. Содержание работы изложено на 137 страницах. Список литературы содержит 71 наименование.

Публикации. Основные результаты по теме диссертации изложены в 9 публикациях [1–9], 7 публикаций входят в библиографическую базу Scopus [3–9].

Содержание работы

Во **введении** обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, приводится обзор научной литературы по изучаемой проблеме, формулируется цель, ставятся задачи работы, сформулированы научная новизна и практическая значимость представляемой работы.

Первая глава посвящена задаче построения глобальных дескрипторов изображения. В **разделе 1.1** приводится обзор существующих подходов к этой задаче, а также отмечается, что все из них формируют дескрипторы в режиме обучения без учителя. В связи с этим в дальнейших разделах автором исследуются возможные варианты построения дескрипторов с использованием многослойных сверточных нейронных сетей, настройка параметров которых осуществлялась в режиме обучения с учителем. В **разделе 1.2** исследуются различные варианты построения дескрипторов на основе выходов полносвязных слоев нейросети. Путем экспериментов на нескольких коллекциях проведено сравнение дескрипторов, построенных с использованием слоев различной глубины, определен слой, формирующий наиболее семантически значимые дескрипторы. Исследована возможность сжатия таких дескрипторов с помощью метода главных компонент, экспериментально показано, что сжатие до 128 – 256 компонент практически не приводит к потерям в качестве поиска. Наконец, описана процедура адаптации дескрипторов под конкретную поисковую задачу, которая заключается в дообучении нейросети на изображениях, семантика которых близка к семантике изображений

из поисковой коллекции. Экспериментально показано, что такая адаптация позволяет значительно повысить семантическую значимость формируемых дескрипторов.

Существенным недостатком формирования дескрипторов с помощью полносвязных слоев нейросети является то, что они могут быть вычислены лишь для квадратных изображений фиксированного небольшого размера. Поэтому в разделе 1.3 автор исследует возможность формирования дескрипторов на основе сверточных слоев, выходы которых могут быть вычислены для изображений любого размера и любой формы. Выходом сверточного слоя нейросети является трехмерный массив и автор сравнивает различные способы агрегации значений в этом массиве для получения вектора фиксированной длины. Автор экспериментально показывает, что дескрипторы, полученные простым поэлементным суммированием, являются более семантически значимыми по сравнению с дескрипторами, полученными более сложными схемами агрегации, предложенными в литературе. Также показано, что такие дескрипторы могут быть сжаты методом главных компонент, что делает их привлекательными в использовании в задачах с большими поисковыми коллекциями.

Вторая глава посвящена исследованию методов компрессии дескрипторов изображений. В разделе 2.1 приведен обзор существующих подходов к задаче компрессии, в частности, описан метод мультиквантизации, который является передовым из существующих методов и обеспечивает наименьшую ошибку компрессии для заданной степени сжатия. Этот метод разбивает каждый дескриптор на несколько подвекторов, соответствующих ортогональным подпространствам, и затем кодирует каждый из полученных подвекторов векторной квантизацией. Словарь квантизации для каждого из подпространств обучается независимо путем кластеризации методом к-средних. В этом же разделе сформулирован основной недостаток метода мультиквантизации, который ограничивает его использование на практике. Независимо формируя словари для каждого из подпространств, метод делает неявное предположение о том, что распределения подвекторов из различных подпространств взаимно независимы. Для реальных данных это предположение часто не выполняется, поэтому качество сжатия методом мультиквантизации может быть невысоким.

В разделе 2.2 автором предлагается метод Аддитивной квантизации, который является обобщением метода мультиквантизации, а также избегает предположения о независимости распределений. Пусть кодируемые векторы имеют размерность D . Параметрами метода Аддитивной квантизации являются M множеств $C^1, \dots, C^M \subset \mathbf{R}^D$, которые называются *словарями*. Каждый словарь содержит в точности K векторов (слов). Далее m -ый словарь обозначается как C^m , и k -ое слово в m -ом словаре как $c^m(k)$. Метод Аддитивной квантизации кодирует вектор $x \in \mathbf{R}^D$ как сумму M слов (по одному слову из каждого словаря). Формально вектор кодируется кортежем длины M , состоящим из номеров слов $[i_1, i_2, \dots, i_M]$, где каждый номер находится в границах от 1 до K .

$$x \approx \sum_{m=1}^M c^m(i_m), \quad i_m \in 1..K \quad (1)$$

Схема работы методов мультиквантизации и Аддитивной квантизации представлена на рисунке 2.

В этом же разделе автор описывает процедуру кодирования вектора x при имеющихся словарях $C^1, \dots, C^M \subset \mathbf{R}^D$, которая заключается в нахождении кода, минимизирующего ошибку компрессии:

$$\min_{i_1, i_2, \dots, i_M} \left\| x - \sum_{m=1}^M c^m(i_m) \right\|^2 \quad (2)$$

Полученная задача минимизации оказывается NP-сложной, поэтому она решается приближенно методом лучевого поиска. Несмотря на отсутствие гарантий оптимальности кодирования, ошибка сжатия в среднем оказывается меньше, чем ошибка полученная мультиквантизацией, что экспериментально продемонстрировано на нескольких коллекциях дескрипторов. Недостатком оптимизации лучевым поиском является то, что вычислительная сложность растет кубически по длине кода M . Поэтому на практике метод Аддитивной квантизации может быть применим только к кодам небольшой длины до 10 байт.

Наконец, автор описывает процедуру обучения словарей Аддитивной квантизации, которая заключается в решении задачи минимизации суммарной ошибки компрессии:

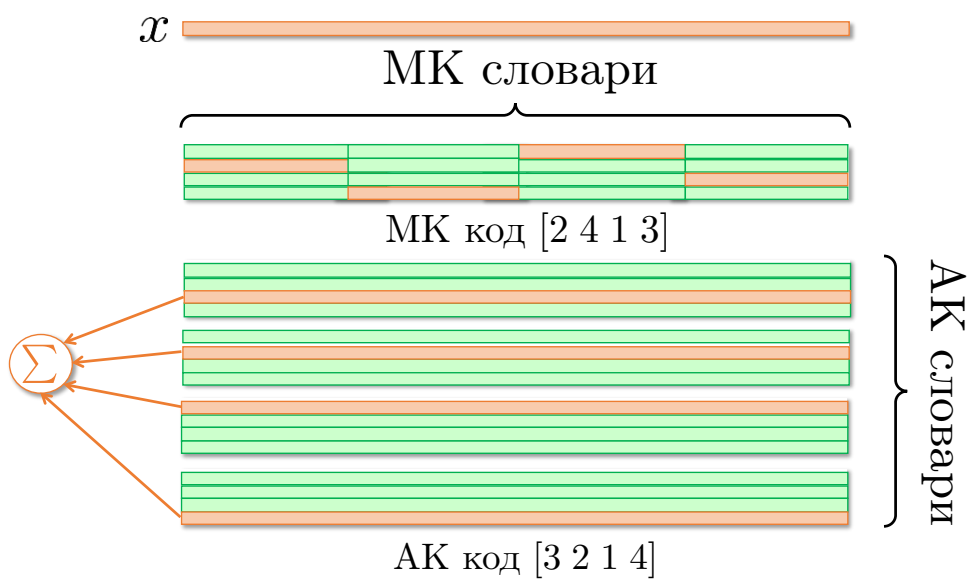


Рис. 2 — Схема работы мультиквантизации (МК) и Аддитивной квантизации (АК) в случае $M=4$ словарей, каждый из которых имеет размер $K=4$. Оба метода кодирования кодируют входной вектор M числами от 1 до K . В случае МК такой код соответствует конкатенации M слов длины D/M . В случае АК, этот код соответствует сумме M слов длины D . При наличии обученных словарей АК достигает меньшей ошибки кодирования входного вектора за счет большего количества настраиваемых параметров.

$$\min_{\substack{C^1, \dots, C^M \subset \mathbf{R}^D \\ |C^m|=K \\ i_j^m \in 1..K}} \sum_{j=1}^n \left\| x_j - \sum_{m=1}^M c^m(i_j^m) \right\|^2 \quad (3)$$

Минимизация осуществляется блочно-координатным спуском, то есть целевая функция по очереди минимизируется по кодам i_j^m (дискретным переменным) и по словам $c^m(\cdot)$ (непрерывным переменным). Предложенный алгоритм оптимизации обобщает стандартный алгоритм k-средних (который эквивалентен случаю $M=1$). Минимизация по дискретным переменным эквивалентна задаче кодирования, решение которой описано выше. Минимизация по непрерывным переменным оказывается задачей наименьших квадратов и решается стандартными пакетами оптимизации.

Практическое преимущество метода Аддитивной квантизации экспериментально продемонстрировано в приложении к задаче поиска ближайшего соседа среди большой коллекции сжатых векторов. Меньшая ошибка кодирования позволяет лучше определять ближайших соседей среди закодированных точек, что приводит к повышению точности поиска.

В разделе 2.3 автор предлагает метод Древесной квантизации, который лишен недостатков Аддитивной квантизации и допускает эффективную процедуру кодирования для любых длин кода. Метод Древесной квантизации также аппроксимирует каждый вектор суммой M слов, но в отличие от Аддитивной квантизации накладывает определенные ограничения на слова, которые определяются структурой “кодирующего дерева” (см. рисунок 3).

Кодирующее дерево $\mathcal{T}$ — граф, являющийся деревом с M вершинами, каждая из которых соответствует некоторому словарию. Далее обозначение $(m, n) \in \mathcal{T}$ для $m \in 1..M$ и $n \in 1..M$ подразумевает, что m -ая и n -ая вершины соединены в дерево ребром. Каждая из D координат в $\mathcal{R}^D$ приписана к единственному ребру в дереве $\mathcal{T}$. Обозначим за $\mathcal{D}_{m,n}$ множество координат, которые приписаны к ребру $(m, n) \in \mathcal{T}$. Так как каждая координата приписана только к одному ребру, эти множества имеют пустое пересечение.

Далее обозначим за $\mathcal{D}_m$ объединение всех множеств размерностей, приписанных к ребрам, инцидентным вершине m , то есть:

$$\mathcal{D}_m = \left\{ \bigcup \mathcal{D}_{m,n} \mid n \in 1..M, (m, n) \in \mathcal{T} \right\}. \quad (4)$$

В модели ДК все слова из m -ого словаря могут иметь ненулевыми только координаты из $\mathcal{D}_m$, то есть:

$$\forall i, d \notin \mathcal{D}_m, c^m(i)[d] = 0. \quad (5)$$

В результате, каждая координата кодируется **двумя** словарями, соответствующими вершинам, инцидентным ребру, к которому она приписана. В этом заключается отличие модели ДК от существующих моделей: в мультиквантизации каждая координата кодируется строго одним словарем, а аддитивная квантизация кодирует каждую координату всеми M словарями. Важным следствием ограничений, накладываемых в модели ДК, является то, что два слова обязательно ортогональны, если они принадлежат словарям, соответствующие вершины которых не соединены ребром в кодирующем дереве.

$$\forall (m, n) \notin \mathcal{T}, \forall i, j : \langle c^m(i), c^n(j) \rangle = 0. \quad (6)$$

Данные ограничения на слова из различных словарей позволяют автору построить эффективную процедуру кодирования с использованием динамического программирования, которая гарантированно находит глобально оптимальный код. Автор также описывает решение задачи обучения кодирующего дерева и словарей Древесной квантизации для конкретной коллекции векторов. Структура дерева и значения слов в каждом словаре являются результатом решения задачи целочисленного программирования, которая может быть решена современными пакетами оптимизации за несколько минут.

В этом же разделе автором доказан следующий теоретический результат о том, что ошибка компрессии некоторого множества векторов методом Древесной квантизацией гарантированно не превосходит ошибки компрессии этих векторов мультиквантизацией:

Утверждение: Пусть $X = \{x_1, \dots, x_L\}$ — множество векторов, пусть $C^1, \dots, C^M$, $|C^m| = K$ — множество словарей мультиквантизации, обученных на этом множестве, и пусть $\{i_j^m\}_{j=1 \dots L}^{m=1 \dots M}$ — коды векторов. Тогда есть возможность обучить словари Древесной квантизации для тех же X , M , K , такие что ошибка аппроксимации множества X с этими словарями не превышает ошибку мультиквантизации со словарями $C^1, \dots, C^M$.

Другими словами, можно гарантировать, что Древесная квантизация кодирует с не меньшей точностью, чем мультиквантизация.

Экспериментально показано, что Древесная квантизация достигает более высокого качества кодирования по сравнению с мультиквантизацией. С другой стороны, Древесная квантизация существенно обходит по скорости кодирования Аддитивную квантизацию. В целом, метод Древесной квантизации является практическим компромиссом между малой ошибкой сжатия и высокой скоростью кодирования. С теоретической точки зрения модель Древесной квантизации также более предпочтительна, так как гарантирует оптимальность кодирования для данных словарей, а также гарантированно не увеличивает ошибку мультиквантизации.

Третья глава посвящена исследованию эффективных методов поиска ближайших соседей среди коллекций из миллиардов векторов. В **разделе 3.1** рассмотрены существующие подходы к решению задачи поиска ближайших соседей на миллиардном масштабе. Единственный существующий метод, способный с приемлемыми точностью и скоростью искать ближайших соседей, основан на идее инвертированного индекса, который разбивает пространство поиска на большое число непересекающихся регионов. Как правило, разбиение осуществляется кластеризацией методом к-средних, а регионы соответствуют полученным клеткам Вороного. Во время поиска для вектора-запроса определяется небольшое число ближайших к нему регионов, и векторы из этих регионов являются кандидатами, расстояние до которых вычисляется затем полным перебором. Таким образом, для общей эффективности поиска должно выполняться два условия: 1) необходимо уметь быстро находить ближайшие регионы; 2) необходимо, чтобы число векторов-кандидатов, участвующих в полном переборе было невелико. Инвертированный индекс не способен одновременно выполнить оба условия эффективности, так как при малом числе регионов каждый из них содержит большое число кандидатов, и их перебор занимает существенное время. С другой стороны, если число регионов велико, то вычислительно дорогим становится определение регионов, ближайших к конкретному запросу.

В **разделе 3.2** автор предлагает новую структуру данных для поиска — инвертированный мультииндекс, который обобщает идею инвертированного индекса и позволяет одновременно выполнить оба условия эффектив-

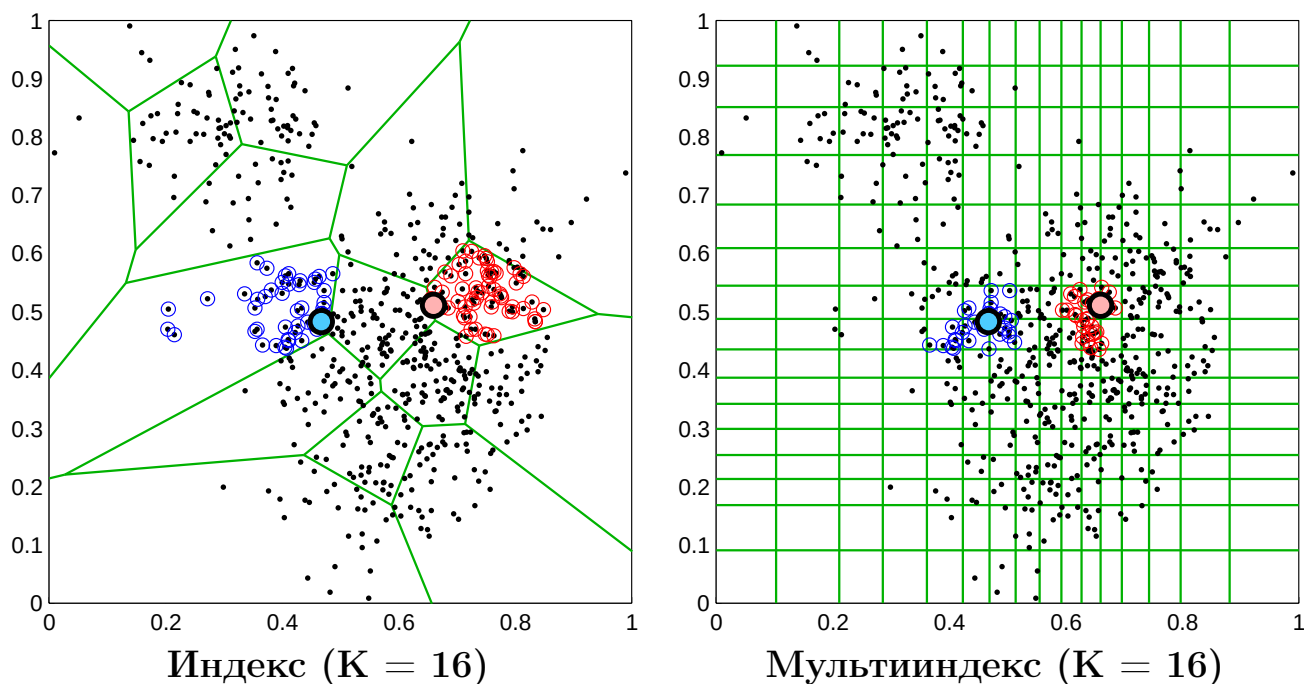


Рис. 4 — Индексация множества из 600 точек (обозначены черным), распределенных неравномерно внутри двумерного квадрата со стороной единица. **Слева** — инвертированный индекс, основанный на стандартной квантизации (словарь имеет 16 двумерных слов; границы регионов обозначены зеленым). **Справа** — инвертированный мультииндекс (каждый из двух словарей имеет 16 одномерных слов). Приведено два примера запросов, обозначенных синим и красным кругами. Списки кандидатов, образованные инвертированным индексом (слева) содержат 45 и 64 элементов соответственно (обведены кругом). Важно отметить, что когда запрос лежит около границы региона (что происходит очень часто в пространствах большой размерности), получившийся список кандидатов может не содержать многих близких точек. Также инвертированный индекс не имеет возможности возвращать список заранее установленной небольшой длины (например, 30 кандидатов). Для одних и тех же запросов, списки кандидатов как минимум длины 30 были сформированы инвертированным мультииндексом, и в этом случае списки содержали 31 и 32 элемента. Даже такие небольшие списки требуют обхода нескольких регионов из разбиения, и получившиеся списки кандидатов гораздо более надежны. В пространствах большой размерности способность обходить регионы, окружающие область запроса с разных направлений, приводит к значительному увеличению точности поиска ближайших соседей.

ности. Для поисковой коллекции размерности D мультииндекс строит два разбиения: 1) разбиение подпространства размерности $D/2$, соответствующего первым $D/2$ размерностям, на регионы с центроидами $C_1^1, \dots, C_K^1 \subset \mathbf{R}^{\frac{D}{2}}$; 2) разбиение подпространства размерности $D/2$, соответствующего вторым $D/2$ размерностям, на регионы с центроидами $C_1^2, \dots, C_K^2 \subset \mathbf{R}^{\frac{D}{2}}$. Два разбиения формируются независимо путем кластеризации методом k -средних и содержат по K регионов размерности $D/2$. Такие разбиения подпространств неявно определяют разбиение исходного пространства поиска на K^2 регионов с центроидами-конкатенациями вида $[C_i^1 \ C_j^2]$, $i, j = 1, \dots, K$. Примеры разбиений, образованных инвертированным индексом и мультииндексом, продемонстрированы на рисунке 4.

В этом же разделе автор описывает эффективную процедуру нахождения ближайших регионов для конкретного запроса и приводит оценки алгоритмической сложности основных операций этой процедуры. Наконец, для коллекции из миллиарда 128-мерных векторов экспериментально продемонстрировано, что по сравнению с инвертированным индексом мультииндекс формирует существенно меньшее число векторов-кандидатов, что позволяет ускорить поиск в 5 — 10 раз без потери точности.

В разделе 3.3 автор показывает, что инвертированный мультииндекс не является оптимальной структурой данных для поиска по миллиардам дескрипторов, полученных с помощью глубоких нейросетей. Причина неоптимальности заключается в том, что разбиения подпространств на регионы в мультииндексе осуществляются независимо, а это неявно предполагает независимость распределений подвекторов в соответствующих подпространствах. Для нейросетевых дескрипторов такое предположение неверно, поэтому автор предлагает еще одну структуру данных: неортогональный инвертированный мультииндекс, которая не предполагает отсутствия корреляций между подпространствами. Неортогональный инвертированный мультииндекс также формирует разбиение на K^2 регионов с центроидами вида $S_i + T_j$, $i, j = 1, \dots, K$, где $S_1, \dots, S_K \subset \mathbf{R}^D$ и $T_1, \dots, T_K \subset \mathbf{R}^D$ — словари. Описаны процедуры эффективного нахождения ближайших регионов и обучения словарей. Экспериментально показано, что для заданного бюджета по времени неортогональный инвертированный мультииндекс позволяет

значительно повысить качество поиска ближайших соседей в коллекции из миллиарда нейросетевых дескрипторов.

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

1. В главе 1 предложено два метода построения дескрипторов изображения на основе глубоких сверточных сетей для задачи визуального поиска: нейросетевые дескрипторы с полносвязных слоев и СПОК дескрипторы. Описан способ адаптации дескрипторов к конкретной поисковой коллекции путем дообучения нейросети, детально представлен протокол сбора коллекции для дообучения. Также исследован вопрос оптимального сжатия нейросетевых дескрипторов, продемонстрировано, что дескрипторы с полносвязных слоев сжимаются методом главных компонент практически без потерь в качестве, в то время как СПОК дескрипторы необходимо сжимать с использованием операции обесцвечивания. На большом количестве тестовых коллекций экспериментально показано, что предложенные нейросетевые дескрипторы более компактны и значительно превосходят предшествующие подходы по точности поиска.
2. В главе 2 предложены два метода сжатия векторов высокой размерности: Аддитивная квантизация и Древесная квантизация, основанные на новой модели аппроксимации векторов с помощью суммы слов. Для обеих моделей описаны процедуры обучения параметров, кодирования и эффективного вычисления расстояний и скалярных произведений. Преимущество модели Аддитивной квантизации состоит в отсутствии каких-либо ограничений на параметры слов, что приводит к наивысшему качеству кодирования в случае коротких кодов. Для длинных кодов кодирование Аддитивной квантизацией вычислительно неэффективно. Модель Древесной квантизации ограничивает возможные значения слов, но допускает эффективное обучение для кодов любой длины. Доказана теорема о том, что ошибка аппроксимации предложенных методов не превосходит ошибки аппроксимации существующих подходов мультиквантизации. Экспериментально показано, что предложенные методы значи-

тельно снижают потери информации при сжатии данных различной природы из прикладных задач.

3. В главе 3 разработаны две структуры данных для эффективного поиска ближайших соседей в поисковых коллекциях из миллиардов векторов: инвертированный мультииндекс и неортогональный инвертированный мультииндекс. Демонстрируется, каким образом мультииндекс разбивает пространство поиска на очень большое число регионов и в то же время позволяет эффективно находить центроиды регионов, ближайшие к запросу. Описано, каким образом структура мультииндекса может быть использована для поиска ближайших соседей и детектирования изображений-дубликатов. Приводятся теоретические оценки алгоритмической сложности операций в мультииндексе. Экспериментально показано, что инвертированный мультииндекс является единственной структурой данных, позволяющей искать ближайших соседей среди миллиардов векторов за несколько миллисекунд.

Публикации автора по теме диссертации

1. *Бабенко Артем, Лемпицкий Виктор*. Эффективный алгоритм поиска ближайших соседей при больших объемах поисковой базы // Труды 54-ой научной конференции МФТИ. Инновации и высокие технологии. — Долгопрудный: 2011. — С. 16–17.
2. *Бабенко Артем, Лемпицкий Виктор*. Компактное кодирование векторов большой размерности для быстрого поиска ближайших соседей // Труды 56-ой научной конференции МФТИ. Инновации и высокие технологии. — Долгопрудный: 2013. — С. 16–17.
3. *Babenko Artem, Lempitsky Victor*. The inverted multi-index // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. — Providence, USA: 2012. — Pp. 3069–3076.
4. Neural Codes for Image Retrieval / Artem Babenko, Anton Slesarev, Alexander Chigorin, Victor Lempitsky // Lecture Notes in Computer Science. Pro-

ceedings of the 13th European Conference on Computer Vision (ECCV 2014), Springer. — Zurich, Switzerland: 2014. — Pp. 584–599.

5. *Babenko Artem, Lempitsky Victor*. Tree quantization for large-scale similarity search and classification // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. — Boston, USA: 2015. — Pp. 4240–4248.
6. *Babenko Artem, Lempitsky Victor*. Additive Quantization for Extreme Vector Compression // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. — Columbus, USA: 2014. — Pp. 931–938.
7. *Babenko Artem, Lempitsky Victor*. Aggregating Deep Convolutional Features for Image Retrieval // Proceedings of the IEEE International Conference on Computer Vision. — Santiago, Chile: 2015. — Pp. 1269–1277.
8. *Babenko Artem, Lempitsky Victor*. The Inverted Multi-Index // *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*. — 2015. — Vol. 37, no. 6. — Pp. 1247–1260.
9. *Babenko Artem, Lempitsky Victor*. Efficient Indexing of Billion-Scale Datasets of Deep Descriptors // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. — Las Vegas, USA: 2016. — Pp. 2055–2063.