

Анализ изображений и видео

1 Введение в компьютерное зрение

1.1 Введение

Наверное, сложно переоценить значимость зрительного восприятия для человека. Зрение необходимо для выполнения практически всех задач в нашей жизни, как повседневных, так и профессиональных. Поэтому, неудивительно, что задача компьютерного зрения столь популярна и востребована. Способность компьютера "видеть" "анализировать" визуальную информацию - один из самых распространённых примеров искусственного интеллекта.

Поскольку человек так хорошо и эффективно воспринимает и обрабатывает визуальную информацию, данные об окружающем нас мире мы часто регистрируем и сохраняем в виде цифровых изображений и видео. Сегодня камера есть в каждом мобильном телефоне, камеры встраиваются в бытовую и профессиональную технику. Мы можем снимать под водой, с воздуха, из космоса. Камеры вводят в человеческое тело, прикрепляют к диким животным. Объёмы мультимедиа информации растут с каждой секундой. Чтобы с этой информацией сделать что-то полезное, нужно научить компьютер обрабатывать и анализировать видео-данные, решать задачи, специфичные для той или иной области.

Давайте посмотрим на некоторые примеры областей применения компьютерного зрения.

- Диагностика опухоли головного мозга по МРТ-снимкам (изображениям, полученным при помощи магнитно-резонансной томографии).
- Автоматическое обнаружение подозрительных объектов с камер видеонаблюдения.
- Большой спектр задач для автономного управления автомобилем: распознавание разметки на дороге, дорожных знаков, других участников движения, и многие другие.
- Анализ спортивных матчей - например, распознавание точной позиции мяча.
- Биометрические системы, как, например, разблокировка телефона по лицу владельца.

- Перевод вывесок, указателей, меню в ресторане с иностранного языка.
- Визуальные системы контроля качества - одно из самых ранних и распространённых применений систем компьютерного зрения. Компьютер анализирует все ли составляющие детали на нужном месте.
- Ещё один пример - автоматизированные магазины Amazon Go, без касс и продавцов. При входе каждый покупатель регистрируется при помощи приложения на мобильном телефоне, и дальше камеры следят за ним по всему магазину, отмечая, что данный покупатель берет с полки и складывает себе в сумку.
- И последний пример на этом слайде - применение компьютерного зрения в сельском хозяйстве. Роботы и дроны на полях - реальность сегодняшнего дня. Роботы определяют степень зрелости урожая, предсказывают его объёмы, распознают наличие заболеваний, используются для автоматической прополки, сбора урожая.

Это всего лишь несколько примеров использования возможностей компьютерного зрения. Мы вернёмся к более подробному обсуждению этих и других приложений чуть позже. А сейчас давайте определим, что же такое компьютерное зрение.

Компьютерное зрение — это область науки, которая занимается алгоритмами и системами фиксации и анализа изображений и извлечения информации из видео-данных. Или, иными словами, машинами, которые могут видеть мир и воспринимать его так же, как человек. По аналогии со зрительной системой человека, система компьютерного зрения состоит из сенсора, фиксирующего изображение, и устройства извлечения информации из изображения. Камеры - это глаза системы компьютерного зрения, а компьютер и алгоритмы анализа и обработки изображений - мозг. На входе система компьютерного зрения, как и зрительная система человека, фиксирует изображение, а на выходе должна быть способна извлечь семантическую информацию, интерпретировать изображение. Например, понять, что это - изображение сада, что какая-то часть изображения соответствует мостику, а какая-то - деревьям, цветам. Людям это кажется простой задачей. Но научить этому компьютер не просто.

В данном курсе мы будем уделять основное внимание второй, и на мой взгляд основной, компоненте системы компьютерного зрения - алгоритмам и методам анализа изображений и видео.

В литературе можно встретить разные определения компьютерного зрения. Для примера приведем три.

- Computing properties of the 3D world from one or more digital images (by Trucco and Veri). В этом определении подразумевается, что вне зависимости от того, есть мы или нет, существует какой-то окружающий мир и его изображения, анализируя которые мы хотим что-то о нём понять. Это определение через имитацию зрительной системы человека. У нас есть сенсор — глаза, у нас есть преобразующее устройство

— мозг, и мы воспринимаем мир при помощи анализа тех картинок, которые видим.

- Make useful decision about real physical objects and scenes based on the sensed images by Shapiro). Это определение ближе к робототехнике. Мы хотим принимать решения и делать выводы о реальных объектах вокруг нас на основе изображений, которые уловили сенсоры. К примеру, это определение идеально подходит под описание того, что делает робот-пылесос. Он принимает решение о том, куда ему дальше ехать и какой угол пылесосить на основании того, что он видит.
- The construction of explicit, meaningful decisions of physical objects from images (by Ballard and Brown). Наиболее общее определение из трех. Если опираться на него, мы хотим просто описывать явления и объекты вокруг нас на основе анализа изображений.

Подытоживая, можно сказать, что компьютерное зрение сводится к извлечению значимой информации из цифровых изображений.

Что из себя представляет цифровое изображение? Существует несколько типов цифровых изображений, самым распространенным из которых являются растровые изображения. Растровое черно-белое изображение - это двумерный массив чисел, или матрица пикселей. Каждый элемент массива кодирует интенсивность света в соответствующей пространственной позиции. Ноль соответствует чёрному цвету - у чёрного цвета нулевая интенсивность света. Как по массиву чисел компьютеру определить, что данное изображение - это портрет Эйнштейна? Это непросто.

Итак, задача компьютерного зрения - извлечь значимую информацию из цифрового изображения. Какую значимую информацию можно извлечь из изображения?

- Первое - семантическую информацию. Ответить на вопрос, что изображено на картинке, понять семантику сцены. Распознать и классифицировать объекты, определить их свойства и отношения. Например, распознать, что на фотографии слева изображён мужчина на мотоцикле с двумя собаками. Идентифицировать личность этого мужчины, породу собак, модель мотоцикла.

Помимо семантической информации, важное значение имеет также метрическая информация об объектах. Определение размеров объектов, расстояния между ними и их взаимного положения - важная подзадача компьютерного зрения. Метрическая информация необходима в робототехнике и системах навигации, в задачах 3D моделирования по снимкам.

Область компьютерного зрения тесно связана с многими другими дисциплинами. На сегодняшний день, большинство задач компьютерного зрения решается при помощи алгоритмов машинного обучения. Компьютерное зрение также связано с когнитивной наукой, занимающейся моделированием

когнитивных систем; с нейронауками, занимающимися изучением нейронных процессов. Компьютерное зрение широко используется в робототехнике, система компьютерного зрения - неотъемлемая компонента большинства роботов. Информационный поиск, анализ и синтез речи - тоже связанные дисциплины.

Обработка изображений, анализ изображений и компьютерная графика - очень тесно связанные области. Некоторые относят все три дисциплины к области компьютерного зрения.

В рамках курса мы будем говорить об алгоритмах обработки изображений. Именно они используются, когда мы повышаем контрастность, удаляем цвет или шум, применяем фильтры и т.д. Любые преобразования изображений, когда на входе и выходе алгоритма изображение — это обработка изображений.

Для анализа изображений характерно, имея на входе изображение, на выходе получить некую модель или набор признаков. То есть по исходному изображению вычислить некоторые числовые параметры, которые описывают это изображение. Например, гистограмму распределения уровней серого цвета, или координаты и классы объектов. В анализе изображений как результат мы получаем набор признаков.

Ещё одна смежная область — компьютерная графика, в которой генерируют изображение по модели или набору признаков.

Все это невозможно без использования знаний и алгоритмов из ещё целого ряда областей, таких как распознавание образов и статистике. Также можно сказать, что анализ изображений — это частный случай анализа данных, и под-область искусственного интеллекта. К смежной дисциплине можно отнести и нейропсихологию — для того чтобы понять, какие у нас есть возможности и как устроено восприятие картинок, полезно понимать, как эти задачи решает наш мозг.

Отсчёт истории компьютерное зрение, как области исследований, принято начинать с 60-х годов 20-го века. В качестве первой точки отсчёта часто приводят студенческий летний проект в группе искусственного интеллекта МИТ 1966 года. Первые попытки обнаружения краёв и восстановления 3D-формы для простых объектов, "мира игрушечных кубиков" (Roberts 1965).

В 1970-х годах сформировался основной понятийный аппарат в области обработки изображений. Велись активные исследования алгоритмов обнаружения краёв, были предложены первые алгоритмы сопоставления стерео и анализа оптического потока.

В 80-х годах большое внимание уделялось математическому аппарату анализа изображений. Многие работы используют Марковские случайные поля для формулировки задач компьютерного зрения, что позволяет рассматривать эти задачи как задачи оптимизации. Сформировалась теория уровней представления изображений. Одной из основных вех этого этапа является книга Дэвида Марра «Зрение. Информационный подход к изучению представления и обработки зрительных образов».

В 90-х годах продолжают активно развиваться многие направления компьютерного зрения: определение формы объекта по движению, методы на

основе оптического потока, определение трехмерной формы объекта по изображениям с разных ракурсов, сегментация изображений (алгоритмы *normalized cuts* и *mean shift*). Возможно самым значимым результатом этого десятилетия стало близкое взаимодействие областей анализа изображений и компьютерной графики - наложение текстуры с реального изображения на 3D модель, построение трехмерных моделей объектов по снимкам с разных ракурсов, и автоматическая генерация новых ракурсов этих объектов. Первые примеры использования статистических методов - использование метода главных компонент для распознавания лиц.

В 2000-е годы связь между компьютерным зрением и компьютерной графикой продолжила развиваться: построение панорам по нескольким снимкам, создание HDR изображений примеры такого взаимодействия. Также на это десятилетие приходится существенные сдвиги в алгоритмах распознавания, классификации изображений, информационного поиска. Было предложено большое количество методов извлечения признаков из изображений, которые затем подавались на вход алгоритмам машинного обучения для решения задач распознавания и классификации.

В последнем десятилетии несомненно самой значительной вехой в компьютерном зрении стало использование алгоритмов глубокого обучения (*deep learning*). Рост объемов данных и развитие аппаратного обеспечения позволили использовать эти вычислительно сложные алгоритмы, которые за несколько лет привели к большому прогрессу во всех задачах компьютерного зрения. До методов глубокого обучения системы компьютерного зрения в лучшем случае работали в контролируемой среде, когда можно отрегулировать освещение, камеру, положение объекта, и т.п. В естественных условиях (неконтролируемой среде) был огромный разрыв в точности распознавания. Однако за последние несколько лет удалось добиться большого прогресса благодаря использованию методов глубокого обучения.

Но несмотря на прогресс последних лет, мы до сих пор водим автомобили самостоятельно, и даже к системам видеонаблюдения часто должен быть приставлен человек. Почему? Одна из ключевых проблем — семантический разрыв.

Человек, глядя на картинку, понимает её семантику. Компьютер же понимает цвет пикселей, умеет выделить текстуру и в конечном счёте отличить кирпичную стену от ковра и распознать на фотографии человека, но определить намерения человека по выражению его лица, позе машина ещё может.

Помимо семантического разрыва, есть ещё ряд сложностей, с которыми сталкиваются разработчики систем компьютерного зрения.

- Один и тот же объект может выглядеть очень по-разному при различном освещении
- Изображение может быть очень разным в зависимости от позы модели
- Фоновый шум и маскировка

- Перекрытия
- Внутрикласовая изменчивость
- Ракурс, точка съёмки

1.2 Примеры практического применения

Мы обсудили, что такое компьютерное зрение и почему это сложно. Давайте теперь рассмотрим несколько областей применения анализа изображений и видео чуть более детально.

1.2.1 Медицина

Одной из основных (и самых первых) областей применения анализа изображений является медицина. Первые попытки автоматически анализировать медицинские снимки при помощи компьютера были предприняты на заре компьютерного зрения. И сегодня эта область остаётся одной из самых востребованных.

Медицинские изображения являются важным источником информации о внутренних органах человека. Они активно используются для диагностики, при планировании и проведении операций, для индивидуального моделирования последствий операций.

Существует множество способов получения медицинских снимков. Изображения получают при помощи электромагнитного излучения, рентгена, ультразвука, компьютерной томографии (КТ), магнитно-резонансной томографии (МРТ). Такое многообразие также говорит о том, насколько важной является визуальная информация в медицине. Большинство медицинской техники, которая используется для получения медицинских снимков, обладает встроенными возможностями автоматической обработки и анализа изображений.

Компьютерный анализ медицинских изображений применим буквально во всех областях медицины. Обнаружение переломов на рентгеновских снимках костей, анализ снимков грудной клетки для диагностики лёгочных заболеваний (например, пневмонии), анализ кровеносных сосудов при помощи ангиографии, обнаружение раковых клеток, диагностика мозга, моделирование зубов. Широкий круг задач в анализе медицинских изображений связан с дерматологией. Список можно продолжать очень долго.

На слайде примеры рентгеновских изображений.

- Рентген кисти руки.
- Рентген грудной клетки.
- Ангиограмма аорты. Для получения ангиограммы в артерию или вену вводят катетер, который продвигается в вдоль сосуда, пока не достигнет обследуемой зоны. Затем через катетер впрыскивается контрастное вещество, хорошо поглощающее рентгеновские лучи. Благодаря

этому усиливается контраст рентгеновского изображения, что позволяет лучше определять аномалии кровоснабжения или места закупорки сосудов.

- Компьютерная томограмма головы.

Какого типа алгоритмы обычно используются для анализа медицинских снимков? По снимкам можно выявлять аномалии — отличается ли снимок данного пациента от снимка здорового человека?

При помощи алгоритмов классификации можно проводить диагностику заболеваний. Если у вас есть база снимков пациентов и известно, что первая аномалия встречается у здоровых людей, а вторая означает, что человек болен раком, то, основываясь на подобии изображений, можно помочь врачам с диагностикой заболеваний.

Анализ изображений используется для индивидуального моделирования организма и предсказания последствий лечения. Хотя мы все и похожи, каждый организм устроен индивидуально. Например, при помощи моделирования зубов пациент может увидеть, как именно его зубы будут меняться от недели к неделе, если поставить брекеты. Другой пример - моделирование последствий операции шунтирования кровеносного сосуда. Если человеку требуется соединить разорванный сосуд шунтом, то определить, где его ставить можно смоделировав по снимку кровеносную систему именно этого пациента и «вставив» шунт в этой модели. Так мы получим возможность увидеть, как изменится кровоток, и предсказать, как пациент будет себя чувствовать при разных вариантах проведения операции.

Анализ изображений незаменим для роботизированной хирургии, для неинвазивных и малоинвазивных операций, которые все больше и больше заменяют открытые операции. Например, операции на мозге без вскрытия черепа - удаление опухолей при помощи сфокусированного ультразвука. Или различные операции с использованием эндоскопических инструментов и дистанционного управления этими инструментами. При таких операциях инструменты вводятся в тело через кожу или через анатомические отверстия. Операция проводится при непрямом наблюдении внутренних органов через инструменты, введенные в тело. И алгоритмы обработки и анализа изображений помогают хирургам видеть и оперировать.

1.2.2 Промышленность

Ещё одна очень востребованная область применения анализа изображений - системы контроля качества на производстве. Для данной прикладной задачи часто используют изображения, регистрируемые в видимом диапазоне (т.е. при помощи обычных оптических камер), и изображения, получаемые при помощи электронного микроскопа.

Компьютерное зрение используется для автоматического контроля выпускаемой продукции давно. Все примеры на слайде из книги 1992 года издания.

- Контроль наличия всех компонент на плате контроллера дискового CD-ROM . Чёрный квадрат в правой верхней части демонстрирует отсутствие микросхемы.
- Контроль наличия всех таблеток в упаковке.
- Контроль за уровнем жидкости в бутылках.
- Контроль за качеством пластмассы - на данном изображении пластмассовая деталь с недопустимым количеством пузырьков воздуха. Обнаружение подобных аномалий составляет важную область промышленного контроля различных материалов, например тканей и дерева.
- Контроль за качеством кукурузных хлопьев - по цвету и наличию подгоревших хлопьев.

Первые алгоритмы работали на основе большого количества эвристик и правил. Алгоритмы создавались под определённое расположение камеры, под определённый продукт. Так что создание и внедрение таких автоматизированных систем контроля качества было оправдано только для больших производственных линий. Сегодня все больше и больше систем контроля качества используют машинное обучение и искусственные нейронные сети.

1.2.3 Системы видеонаблюдения

Ещё одна прикладная область, где необходимы алгоритмы анализа изображений и видео, это системы видеонаблюдения. Сегодня камеры наблюдают за нами повсюду: в аэропортах, на вокзалах, в метро, в магазинах, просто на улицах. При таком объёме видеоданных без автоматизированных систем анализа видеоряда не обойтись. Большинство камер устанавливаются для обеспечения безопасности. Системы видеораспознавания позволяют оперативно распознать скопления людей, идентифицировать и зафиксировать факты их неадекватного движения (драки, падения, хаотичное движение, резкое ускорение, пересечение запрещённой зоны), обнаружить в контролируемой зоне оставленные или исчезнувшие предметы, выявить задымление или открытый огонь. Задачи по обнаружению подозрительных предметов или подозрительного поведения не так просты. Их сложность в том, что часто нельзя заранее дать описание того, что является подозрительным предметом. Невозможно предоставить все возможные примеры аномального поведения. Поэтому часто в таких системах моделируют нормальное поведение, и распознавание аномалий устроено на выявлении отклонений от нормы.

Помимо охранных целей, системы видеонаблюдения часто используются для изучения поведения людей. Например, при помощи установленных камер в магазине можно определить наиболее популярные отделы магазина и использовать эту информацию для более удобной планировки расположения товаров.

Системы видеонаблюдения позволяют автоматически собирать плату за проезд по платному участку дороги, штрафовать за превышение скорости, наблюдать за животным миром. Список задач бесконечен!

1.2.4 Биометрия

Ещё одним направлением, активно использующим компьютерное зрение, являются биометрические системы идентификации личности. Такие системы основаны на измерении уникальных биологических и физиологических характеристик, индивидуальных для каждого человека. Иногда эти характеристики ещё называют биологическим кодом человека.

Преимущество биометрических систем идентификации, по сравнению с традиционными (например, PIN-кодовыми системами или системами доступа по паролю), заключается в том, что идентифицируется не какой-то внешний предмет, принадлежащий человеку, а собственно сам человек. Анализируемые характеристики неразрывно связаны с человеком, их невозможно потерять, передать, забыть и крайне сложно подделать. К тому же эти характеристики практически не подвержены износу и не требуют замены или восстановления.

Среди основных биометрических технологий, работающих на основе компьютерного зрения - распознавание личности по отпечаткам пальцев, геометрии лица, геометрии кисти руки, радужной оболочке или сетчатке глаза, геометрии рисунка вен, по подписи и др.

Первые биометрические системы были достаточно дорогими и требовали значительных вычислительных ресурсов. Сегодня мы разблокируем компьютеры и телефоны по отпечатку пальца или геометрии лица.

Приведу ещё одну нашумевшую в СМИ историю, связанную с задачами биометрии. В Википедии даже есть статья, посвящённая этой истории. В 80-х годах фотограф National Geographic сфотографировал неизвестную афганскую девочку. Фотография стала очень известной после того, как появилась на обложке журнала National Geographic в июне 1985 года. Поскольку личность девочки была неизвестна, фотографию называли просто «Афганской девочкой» (Afghan Girl). Фотографию иногда сравнивают с портретом Моны Лизы кисти Леонардо да Винчи и называют «Афганской Моной Лизой». На протяжении 1990-2000 годов журналист предпринимал несколько попыток узнать имя девочки. Несколько женщин ошибочно объявляли себя «Афганской девочкой»; многие молодые люди называли её своей женой. Через 17 лет после исходной фотографии фотограф нашёл её в Афганистане и её личность была подтверждена при помощи биометрии, показавшей полное соответствие радужной оболочки глаза изображению на фотографии.

1.2.5 Индексирование и поиск изображений

Индексирование и поиск изображений ещё одна важная задача, для решения которой требуются алгоритмы анализа изображений. Коллекции и

архивы изображений бывают разные.

- Персональные. Например, в отпуске человек может сделать пару тысяч фотографий, с которыми потом нужно что-то делать.
- Профессиональные. Они насчитывают миллионы фотографий. Здесь тоже есть необходимость как-то их организовывать, искать, находить то, что требуется.
- Коллекции репродукций. Это тоже миллионы изображений. Сейчас у большого количества музеев есть виртуальные версии, для которых оцифровываются репродукции. Пока утопичная задача — поиск всех репродукций одного и того же автора. Человек по стилю может предположить, что видит, допустим, картины Сальвадора Дали. Было бы здорово, если бы этому научилась и машина.

Как упорядочить все эти коллекции? Какие ещё задачи можно решать? Можно построить систему навигацию по коллекциям, классифицируя их по темам. Отдельно складывать мишек, отдельно слонов, отдельно апельсины — так, чтобы пользователю потом было удобно ориентироваться по этой коллекции.

Ещё одна задача — это поиск дубликатов. В двух тысячах фотографий из отпуска неповторяющихся значительно меньше. Мы любим экспериментировать, снимать с разной выдержкой, фокусным расстоянием и т.д., что в итоге нам даёт большое количество "нечётких" дубликатов, т.е. примерных копий. Кроме того, поиск по дубликатам может помочь обнаружить незаконное использование вашей фотографии, которую вы однажды могли выложить в интернете.

Ещё одна интересная задача — выбор лучшей фотографии. С помощью алгоритма можно понять, какая картинка больше всего понравится пользователю. Например, если это портрет, лицо должно быть освещено, глаза открыты, изображение должно быть чётким и т.д. В современных фотоаппаратах есть такая функция.

Ещё одна задача поиска — создание коллажей, т.е. подбор фотографий, которые будут хорошо смотреться рядом.

1.2.6 Ещё приложения

Есть ещё большое количество прикладных задач, где используется анализ изображений.

- Системы слежения и целенаведения в военной промышленности. Очевидными примерами являются обнаружение вражеских солдат и транспортных средств и управление ракетами. Наиболее совершенные системы управления ракетами посылают ракету в заданную область, вместо конкретной цели, а выбор целей производится, когда ракета достигает заданной области, основываясь на получаемых видеоданных.

- Автономные транспортные средства, как для военного применения, так и для автомобилестроения в целом. Уровень автономности изменяется от полностью автономных (беспилотных) до транспортных средств, где системы, основанные на компьютерном зрении, поддерживают водителя или пилота в различных ситуациях. Полностью автономные транспортные средства используют компьютерное зрение для навигации, то есть для получения информации о месте своего нахождения, для создания карты окружающей обстановки, для обнаружения препятствий. Примерами таких систем могут быть система предупредительной сигнализации о препятствиях на машинах, системы автономной посадки самолётов. Сегодня активно разрабатываются и тестируются системы автономного управления автомобилем.
- Компьютерное зрение также активно используется в киноиндустрии при создании спецэффектов, для совмещения игры живых актёров с компьютерной анимацией. Сегодня создатели фильмов могут записать эмоции настоящего актёра и наложить их на нарисованного персонажа или на лицо другого актёра. Компьютерное зрение сегодня - неотъемлемая часть абсолютного большинства компьютерных игр, приложений виртуальной и дополненной реальности.
- Зондирование земной поверхности с использованием видимого и инфракрасного диапазонов спектра - ещё одна важная область применения алгоритмов анализа изображений. Изображения населённых пунктов часто (и давно) используются для оценки роста численности населения, динамики загрязнения окружающей среды. По изображениям со спутников автоматически составляют карты местности. Наблюдение за погодой и составления прогнозов также является важным применением спутниковых изображений. Обнаружение и предсказание ураганов, обнаружение лесных пожаров, и даже изменение уровня поверхности земли как показатель истощения подземных вод - все это делается сегодня при помощи анализа спутниковых снимков земли.
- Оцифровка печатных документов, обнаружение и распознавание текста - ещё один пример применения алгоритмов анализа изображений. Сегодня существуют приложения, которые позволяют в режиме реального времени переводить указатели и таблички с иностранного языка.
- И, конечно, компьютерное зрение - неотъемлемая компонента системы навигации и управления роботов.

1.3 Задачи и наборы данных

1.3.1 Задачи

Мы рассмотрели большое число различных прикладных задач, в которых используется компьютерное зрение. Как же можно научить компьютер "по-

нимать "изображения, чтобы можно было решать все эти задачи?

Давайте посмотрим на одно видео. Ваша задача - посчитать сколько раз передают мяч друг другу игроки в белом.

Заметили ли вы, что среди игроков прошёл человек, наряженный в костюм медведя? Давайте посмотрим видео ещё раз.

После того, как я вам сказала про медведя, Вы его, конечно заметили. Потому что Вы знали, что искать.

Это видео демонстрирует, что даже зрительная система человека обычно в каждый конкретный момент времени решает какую-то одну поставленную задачу, игнорируя сигналы, которые не относятся к решению этой задачи. Точно также при разработке приложений компьютерного зрения люди обычно решают какую-то одну конкретную поставленную задачу. Это существенно проще, чем пытаться решить абстрактную задачу "научить компьютер понимать изображение".

Какие же конкретные задачи чаще всего решаются в различных системах компьютерного зрения?

- Обнаружение лиц - найти на изображении все человеческие лица.
- Распознавание лиц - определить, какому человеку соответствует данное лицо.
- Идентификация лиц - определить, соответствует ли данное лицо какому-то конкретному человеку.
- Распознавание позы - определить какой позе из набора заранее заданных поз соответствует данная поза человека.
- Выделение областей, однородных по цвету или текстуре.
- Распознавание объектов - определить какому из заранее определённых классов относится данный объект на изображении.
- Семантическая сегментация - выделение областей изображения, относящихся к одному объекту.
- Классификация изображений - определение к какой категории относится изображение или изображённые на нем объекты, без локализации этих объектов на изображении. При помощи классификации изображений можно ответить на такие вопросы об изображении как изображён ли на картинке слон, или самолёт, или закат. Без указания где именно на картинке расположен слон или самолёт.
- Обнаружение и распознавание текста - определить, где на картинке изображён текст и распознать что там написано.

И даже такая общая задача, как семантическое описание изображения может быть решена как совокупность перечисленных ранее более узких задач. Например, давайте посмотрим на эту фотографию. Если я вас попрошу

описать эту фотографию словами, скорее всего я услышу, что на фотографии изображена маленькая девочка, которая ест мороженное. Вы сможете описать во что она одета, где она находится.

можно составить похожее описание изображения автоматически при помощи компьютера, решив такие задачи, как распознавание объектов и лиц, определение пола и возраста человека, выделение однородных по цвету областей, распознавание действия, выделение текстуры.

1.3.2 Наборы данных

Очень важную роль в развитии области компьютерного зрения (как и во многих других областях анализа данных) играют общедоступные наборы данных, собранные для тестирования и сравнения различных алгоритмов, решающих какую-то конкретную задачу. Стандартные наборы данных необходимы для объективного сравнения различных подходов к решению одной и той же задачи. Данные необходимы для разработки алгоритмов, основанных на машинном обучении.

Давайте познакомимся с наиболее известными наборами данных, используемыми научным сообществом для разработки и тестирования все более и более совершенных алгоритмов анализа изображений.

Lena Наверное, самым известным тестовым изображением, использовавшимся для проверки и демонстрации различных алгоритмов обработки изображений, является «Лена». Тестовое изображение представляет собой оцифрованный портрет шведской модели, который в свою очередь является фрагментом разворота из журнала Playboy. Согласно Википедии, в 1973 году Александру Савчуку из Университета Южной Калифорнии для иллюстрации статьи на тему обработки изображений понадобился фотопортрет с хорошим динамическим диапазоном. Савчук отсканировал фрагмент постера из Playboy. Он использовал сканер с разрешением 100 линий на дюйм, и результирующее изображение получилось 512 на 512 точек. Вскоре эта картинка превратилась для индустрии в стандарт де-факто: на нём проверялись и отрабатывались всевозможные приёмы коррекции изображений, оттачивались новые алгоритмы обработки.

MNIST Одним из первых наборов изображений для тестирования алгоритмов машинного обучения является коллекция MNIST (сокращение от «Modified National Institute of Standards and Technology»). Она состоит из образцов рукописного написания цифр. С использованием этого набора данных разрабатывались первые алгоритмы анализа изображений на основе нейронных сетей. Коллекция MNIST содержит 60000 изображений для обучения и 10000 изображений для тестирования размером 28x28 пикселей. Все изображения поделены на 10 классов, по числу цифр. Задача распознавания рукописных цифр на этом наборе данных - это классическая задача классификации. Необходимо по изображению определить к какому

классу оно принадлежит (т.е. какая из 10 цифр изображена). По сегодняшним меркам, и размер коллекции, и разрешение изображений очень небольшой. Но этот набор данных по-прежнему активно используется, в основном при обучении нейронным сетям. Запрограммировать нейронную сеть для обучения и тестирования на этом наборе данных, это как написать "Hello World" на новом языке программирования.

CIFAR CIFAR-10 и CIFAR-100 - тоже популярные наборы данных для тестирования нейронных сетей. По количеству и разрешению изображений они похожи на MNIST, но изображения более разнообразны. CIFAR-10 содержит картинки из 10 классов, таких как самолёт, машина, птица, и тому подобные. CIFAR-100 содержит картинки из 100 классов. Эти наборы данных также широко используются для тестирования алгоритмов классификации изображений.

ImageNet ImageNet - размеченная коллекция изображений, с созданием которой многие связывают начало эпохи глубокого обучения и успех глубоких нейронных сетей. Для обучения глубоких сетей нужно много данных. ImageNet - первый действительно большой набор аннотированных изображений. ImageNet содержит более 14 миллионов изображений (сравните с 60000 в MNIST и CIFAR!), разбитых на более чем 20000 классов.

С использованием этой коллекции данных с 2010 года проводится соревнование ILSVRC (англ. ImageNet Large Scale Visual Recognition Challenge — Кампания по широкомасштабному распознаванию образов в ImageNet), в рамках которого различные программные продукты ежегодно соревнуются в классификации и распознавании объектов и сцен в базе данных ImageNet. Для ILSVRC используется подмножество коллекции ImageNet, состоящее из около 1.2 миллиона изображений и 1000 классов.

PASCAL VOC До появления ImageNet и ILSVRC наиболее известным соревнованием по классификации изображений и обнаружению объектов было PASCAL VOC challenge (Pattern Analysis, Statistical Modelling and Computational Learning Visual Object Classes), проводившееся с 2005 по 2012 год. Коллекций изображений, собранные для этого соревнования, долгое время являлись де-факто стандартом для тестирования различных алгоритмов, особенно для алгоритмов обнаружения объектов и сегментации. PASCAL VOC в отличие от предыдущих обсуждавшихся коллекций содержит разметку на уровне объектов, когда на изображении размечены регионы с объектами.

<http://host.robots.ox.ac.uk/pascal/VOC/databases.html>

<http://host.robots.ox.ac.uk/pascal/VOC/images/tud3c.html>

Caltech Caltech101 и Caltech256 ещё две коллекции изображений для тестирования алгоритмов классификации изображений и распознавания объектов. Коллекции содержат изображения 101 и 256 категорий соответствен-

но. Размер картинок примерно 200x300 пикселей - как в ImageNet и PASCAL VOC. Большинство категорий Caltech101 содержит около 50 изображений, но есть категории с намного большим числом картинок. Например в категориях "самолет" и "лицо" около 800 примеров.

К каждому изображению прилагается аннотация, содержащая координаты рамки в которой находится объект и класс объекта.

http://www.vision.caltech.edu/Image_Datasets/Caltech101/

http://www.vision.caltech.edu/Image_Datasets/Caltech256/

MS COCO Все рассмотренные нами ранее коллекции содержат изображения относительно небольшого размера. В то же время, большинство камер сегодня снимает с более высоким разрешением. И для ряда задач (например, распознавание объектов для автономных транспортных средств) важно уметь работать с изображениями с более высоким разрешением. MS COCO (Microsoft COCO: Common Objects in Context) содержит изображения с разрешением до 1000 пикселей вдоль большей стороны. Также большинство коллекций, о которых мы говорили до этого, не имеет разметки для семантической сегментации, когда для каждого пикселя изображения есть метка о его принадлежности к определённому объекту.

Коллекция MS COCO сегодня является стандартом для сравнения алгоритмов распознавания объектов и семантической сегментации. Она содержит 328000 изображений с объектами 91 категории. Всего в коллекции около 2500000 размеченных объектов.

<http://cocodataset.org/#home>

<https://arxiv.org/pdf/1405.0312.pdf>

Cityscapes Ещё один широко используемый набор изображений высокого разрешения - Cityscapes. Коллекция содержит изображения городских пейзажей, собранных в 50 городах Европы. Размер изображений 1024 x 2048 пикселей. В наборе данных 5000 изображений с детальной аннотацией объектов - на уровне отдельных пикселей, и около 20000 изображений с менее детальной аннотацией.

<https://www.cityscapes-dataset.com/>

<https://arxiv.org/pdf/1604.01685.pdf>

Labeled Faces in the Wild Помимо коллекций, содержащих картинки всевозможных категорий объектов, существуют также наборы данных для распознавания определённых категорий, имеющих особую значимость. Так, существуют размеченные наборы для обнаружения пешеходов (<http://coding-guru.com/popular-pedestrian-detection-datasets/>), распознавания жестов, распознавания дорожных знаков.

Отдельного внимания всегда заслуживала задача обнаружения и распознавания лиц. Было создано большое количество размеченных коллекций для этих задач. Одной из наиболее широко используемых сегодня коллекций является Labeled Faces in the Wild:

<http://vis-www.cs.umass.edu/lfw/>.

Как следует из названия, коллекция содержит изображения лиц в "естественных условиях без контроля над освещением, позой человека и его головы. Это не портретная съёмка на монотонном фоне, а репортажные кадры. Коллекция содержит более 13000 изображений известных людей, собранных в интернете. Около 1680 личностей из коллекции имеют более одной фотографии.

Медицинские изображения Несмотря на то, что анализ медицинских изображений является одной из основных прикладных задач компьютерного зрения, общедоступных коллекций медицинских изображений не так много. В первую очередь это связано с персональным аспектом этих данных. Только в последние годы, когда стало ясно что доступность больших коллекций данных способна значительно ускорить развитие алгоритмов, начали появляться объёмные анонимизированные коллекции медицинских снимков. Первой такой коллекцией, получившей распространение среди групп машинного обучения, была ChestXray14, содержащая около 112,000 снимков грудной клетки с признаками 14 видов болезней. Недавно группа машинного обучения Стэнфорда выпустила ещё одну объёмную коллекцию CheXpert. Эта коллекция содержит 224316 снимков грудной клетки.
<http://academictorrents.com/details/557481faacd824c83fbf57dcf7b6da9383b3235a>
<https://stanfordmlgroup.github.io/competitions/chexpert/>
<https://arxiv.org/abs/1901.07031>

1.4 Основы зрительного восприятия человека

Системы компьютерного зрения, во многом, строятся по аналогии со зрительной системой человека. Если мы хотим научить машину видеть мир и анализировать визуальную информацию так, как это делает человек, то сначала надо разобраться, как же работает зрительная система человека. Как изображение формируется в глазу? Какого рода информацию человек способен извлечь из визуального сигнала? Какими физическими ограничениями обладает система зрительного восприятия человека? Всё это не только интересно, но и важно в практическом плане.

Зрительная система человека обладает удивительными способностями. Мы распознаём людей и предметы в самых различных условиях: с различных ракурсов, при разном освещении, даже если видна только часть объекта. Мы способны с лёгкостью извлечь информацию о размерах предметов, местоположении относительно друг друга, расстояниях между объектами. Наш мозг способен "достроить" картинку, если видна только ее часть.

В то же время, наша зрительная система далеко не идеальна. Мы подвластны иллюзиям, не замечаем многие детали, наше восприятие неоднозначно и субъективно, зависит от нашего "зрительного опыта".

Давайте посмотрим на несколько примеров, иллюстрирующих способности зрительной системы человека.

Как видим мы- внутриклассовая изменчивость объектов Посмотрите на картинки на этом слайде. Для нас не составляет никакого труда понять, что всё это - изображения разных видов ламп, даже если некоторые из этих ламп вы видим впервые в жизни. Лампы могут быть разной формы, разного цвета, разного размера. Но мы всё равно понимаем, что это всё - лампы.

Как видим мы - неоднозначность восприятия А что вы видите на этих изображениях? На самой левой картинке можно увидеть молодую женщину или старуху в платке, на центральном - эскимоса или индейца, на правом - саксофониста или лицо женщины. Наше восприятие неоднозначно.

Несколько фактов о нашем зрительном восприятии - 1 Мозг и наш зрительный опыт играет очень важную роль в том, как мы интерпретируем зрительный сигнал. Мозг способен "достроить" картинку, даже если видна только часть, или добавить семантику, распознать объект, даже когда его там нет. Человек часто даёт не соответствующее действительности объяснение наблюдаемой им картине. Мы все можем узнать "что-то" или "кого-то" в очертании облака.

То, как наш мозг достраивает и интерпретирует картинку очень сильно зависит от нашего зрительного опыта. Наша зрительная система "самообучается" на протяжении всей нашей жизни. Нам проще распознавать изображения, привычные нашему глазу. Например, типичному европейцу сложно различать лица азиатов, а людям выросшим среди африканцев, все белые будут казаться "на одно лицо".

Посмотрите на эту черно-белую картинку. Можете распознать, что это? А если перевернуть? В привычной ориентации существенно проще распознать очертания карты мира.

Несколько фактов о нашем зрительном восприятии - 2 Ещё пара иллюстраций к тезису о том, что мы часто ищем в изображении знакомые образы, и наш мозг "достраивает" картинку. Это примеры известных оптических иллюзий, в которых глаз восполняет несуществующую информацию.

На левом рисунке ясно видны очертания квадрата, вопреки тому факту, что на изображении отсутствуют линии, определяющие такую фигуру. Аналогичный эффект, на этот раз в виде круга, виден на левом рисунке. Всего нескольких линий достаточно для получения иллюзии круга.

Несколько фактов о нашем зрительном восприятии - 3 От того, где мы выросли, зависит, как мы "видим" мир вокруг нас. Например, на этом рисунке ступеньки идут вверх или вниз? Большинство из нас, кто пишет и читает слева направо, ответят, что вверх. А для тех, кто привык читать справа налево, ступеньки будут вести вниз.

Строение человеческого глаза Глаз имеет почти сферическую форму со средним диаметром около 20 мм. По сути, глаз - это оптическая сфера. Свет проникает в глаз через зрачок - отверстие в центре радужной оболочки глаза, преломляется хрусталиком и проецируется на внутреннюю оболочку глаза - сетчатку.

Зрачок может сужаться или расширяться, регулируя тем самым количество попадающего света. Диаметр зрачка может изменяться в пределах от 2 до 8 мм.

Хрусталик, расположенный за радужной оболочкой глаза и зрачком, играет роль оптической линзы. В отличие от жёсткой оптической линзы, оптическая сила хрусталика может изменяться за счёт изменения его формы. Как видно на рисунке, радиус кривизны передней поверхности капсулы хрусталика больше, чем задней. Изменение формы хрусталика происходит путём ослабления или натяжения передней и задней порции волокон ресничного пояса. Для фокусировки зрения на удалённом предмете ресничная мышца расслабляется, хороиidea сжимается, натягивая при этом волокна ресничного пояса, в результате чего хрусталик приобретает уплощенную форму. А для фокусировки на близкорасположенном предмете ресничная мышца сокращается, что приводит к округлению хрусталика и увеличению его преломляющей способности. При изменении преломляющей способности хрусталика с минимального значения до максимального, его фокусное расстояние изменяется от 17 мм до 14 мм. При рассмотрении предметов на расстоянии больше 3 метров преломляющая способность хрусталика минимальна, при разглядывании предметов вблизи, преломляющая способность хрусталика максимальна.

При правильной оптической фокусировке глаза свет от наружного объекта проецируется в виде изображения на сетчатку. По внутренней поверхности сетчатки распределены дискретные светочувствительные клетки (рецепторы), которые регистрируют изображение. Эти рецепторы бывают двух видов - колбочки и палочки. Колбочки обладают высокой чувствительностью к спектральным составляющим света и обеспечивают фотопическое зрение - зрение в ярком свете. В глазу насчитывается от 6 до 7 миллионов колбочек. Все они располагаются преимущественно в центральной области сетчатки, называемой жёлтым пятном. В центре жёлтого пятна имеется так называемая центральная ямка - область наибольшей остроты зрения. Человек различает мелкие детали изображения в основном благодаря колбочкам. Каждая колбочка соединена с отдельным нервным окончанием. Наружные мышцы глаза обеспечивают вращение глазного яблока так, чтобы изображение интересующего объекта попало в область жёлтого пятна.

Палочек в глазу намного больше, чем колбочек - от 75 до 50 миллионов. В отличие от колбочек, палочки распределены по всей поверхности сетчатки, и к одному нервному окончанию присоединено сразу несколько палочек (в среднем, около 10). Это уменьшает их способности различать детали изображения. Палочки позволяют сформировать общую картину всего поля зрения. Они наиболее чувствительны при низких уровнях освещённости и не участвуют в обеспечении функции цветного зрения. Поэтому

предметы, имеющие яркую окраску при ярком освещении, выглядят как лишённые цветов образы при сумеречном освещении, так как при сумеречном освещении возбуждаются только палочки. Это явление известно как скотопическое (или сумеречное) зрение.

Яркостная адаптация и контрастная чувствительность - 1 Знание о способности глаза различать разные уровни яркости имеет важное практическое значение. Оно поможет выбрать правильную схему дискретизации уровня яркости для представления цифровых изображений.

Зрительная система человека способна адаптироваться к огромному диапазону значений яркости, порядка 10^{10} - от порога чувствительности скотопического зрения до предела ослепляющего блеска. Эксперименты также показывают, что субъективная яркость (т.е. яркость, как она воспринимается человеком) является логарифмической функцией от физической яркости света, попадающего в глаз. На слайде изображен график этой зависимости субъективной яркости от истинной яркости. Длинная сплошная кривая представляет диапазон яркостей, в котором способна адаптироваться зрительная система человека. При использовании одного фотопического зрения этот диапазон составляет около 10^6 . При низких уровнях яркости происходит постепенный переход от фотопического к скотопическому зрению.

Яркостная адаптация и контрастная чувствительность- 2 Но важно понимать, что это динамический диапазон - зрительная система человека не способна работать во всем этом диапазоне одновременно. Она охватывает такой большой диапазон за счёт изменения общей чувствительности. Это явление известно как яркостная адаптация. Общий диапазон одновременно различаемых уровней яркости относительно мал по сравнению со всем диапазоном адаптации. Для любого данного набора внешних условий текущий уровень чувствительности зрительной системы, называемый уровнем яркостной адаптации, соответствует некоторой яркости, например, точке B_a на графике. Короткая кривая, пересекающая основной график, представляет диапазон субъективной яркости, которую способен воспринимать глаз при адаптации к указанному уровню. Этот диапазон достаточно ограничен: все уровни яркости ниже B_b субъективно воспринимаются зрением как чёрные и неразличимы. Верхняя пунктирная часть этой кривой реально не ограничена, но теряет смысл при большой длине, поскольку при повышении яркости просто повышается уровень адаптации B_a .

Способность зрения различать изменения яркости при данном уровне адаптации также представляет значительный интерес. Классический эксперимент для определения способности зрительной системы человека различать разные уровни яркости состоит в следующем. Испытуемого просят смотреть на большой плоский равномерно освещённый экран. Экран должен быть такого размера, чтобы занимать все поле зрения испытуемого. Яркость экрана (обозначенная I на слайде) можно регулировать. На

это равномерно освещённый экран накладывается добавочная яркость δI в форме кратковременной вспышки в области круглой формы, расположенной в центре экрана, как изображено на слайде.

Малые приращения δI неразличимы для испытуемого - он не замечает вспышек. По мере увеличения δI в какой-то момент испытуемый начнёт замечать вспышки, т.е. будет воспринимать изменение яркости. Величина $\delta I_c/I$, где δI_c - величина приращения яркости, различимая в 50% случаев на фоне яркости I , называется отношением Вебера. Малое значение отношения Вебера означает, что различаются очень малые относительные изменения яркости, т.е. имеет место высокая контрастная чувствительность. Большое значение отношения Вебера означает, что требуется большое относительное изменение яркости, чтобы его можно было заметить. Это говорит о низкой контрастной чувствительности.

График зависимости логарифма отношения Вебера от логарифма фоновой яркости ($\log(\delta I_c/I)$ от $\log(I)$) изображён на слайде. Эта кривая показывает, что низкая контрастная чувствительность (т.е. большое отношение Вебера) наблюдается при малых уровнях яркости. Контрастная чувствительность заметно возрастает при увеличении фоновой яркости. Наличие двух частей кривой отражает тот факт, что при малых уровнях яркости зрение осуществляется благодаря палочкам, а при больших уровнях яркости - за счёт колбочек.

Если поддерживать фоновую яркость постоянной, а яркость добавочного источника варьировать не вспышками, а ступенчатым изменением яркости от неотличимого до заметного всегда, то типичный наблюдатель способен различить всего 10-20 ступеней яркости. Грубо говоря, это означает, что в произвольной точке монохромного изображения, человек способен различить всего 10-20 уровней яркости. Но это не означает, что всё изображение может быть представлено таким небольшим числом градаций яркости. По мере движения глаза по изображению меняется среднее значение яркости фона, что позволяет нам обнаруживать различные множества относительных изменений яркости для каждого нового уровня адаптации. Следствием этого является способность глаза различать яркости в намного более широком общем диапазоне, чем 20 уровней.

Воспринимаемая яркость, как функция истинной яркости: полосы Маха Воспринимаемая яркость не является простой функцией истинной яркости. Известны два явления, подтверждающие это утверждение. Первое основывается на том факте, что вблизи границ соседних областей с отличающимися, но постоянными яркостями зрение человека склонно "подчёркивать" яркостные перепады. Посмотрите на картинку на слайде. Нам кажется, что яркость полос не постоянна, что вблизи границы с более светлой полосой яркость становится ниже, а вблизи границы с более тёмной полосой яркость становится выше. Хотя на самом деле яркость каждой из полос постоянна. Эти полосы с кажущимися изменениями яркости на краях называются полосами Маха в честь Эрнста Маха, впервые описавшего этот

феномен в 1865г.

Воспринимаемая яркость, как функция истинной яркости: одновременный контраст Второе явление, называемое одновременным контрастом, связано с тем фактом, что воспринимаемая яркость некоторой области определяется не только ее собственной истинной яркостью, но и яркостью соседних областей. В верхней части слайда все три внутренних квадрата имеют одинаковую яркость. Но зрительно левый внутренний квадрат воспринимается светлее, чем правый - потому что фон левого квадрата темнее, чем фон правого. Чем более светлый фон, тем темнее кажется объект на этом фоне.

Тот же эффект наблюдается и на картинке в нижней части слайда. Обе тонкие полосы имеют одинаковую яркость, но нам кажется, что левая полоса светлее правой.

Ещё одним примером является лист белой бумаги, который кажется белым, когда он лежит на столе, но может показаться совершенно чёрным, если им закрывать глаза, глядя на яркое небо.

Иллюзия с тенью на шахматной доске И ещё одна иллюзия, в основе которой также лежит феномен одновременного контраста. Зрительно квадрат А нам кажется темнее квадрата В.

Иллюзия с тенью на шахматной доске-2 Хотя на самом деле, оба квадрата, А и В, абсолютно одинаковой яркости.

1.5 Цвет

Мы поговорили немного о яркости и её восприятии человеком. Давайте теперь обратимся к цвету. Что есть цвет? Как наша зрительная система распознает разные цвета?

Как и в случае с яркостью, воспринимаемый нами цвет зависит не только от физических свойств освещения и объектов. Строго говоря, воспринимаемый нами цвет - это результат взаимодействия между световыми волнами и нашей зрительной системы.

Спектр Хотя процесс восприятия и интерпретации цвета человеческим мозгом представляет собой не до конца исследованное психофизиологическое явление, физическая природа цвета может быть точно описана.

Ещё в 1666г. сэр Исаак Ньютон обнаружил, что при прохождении луча солнечного света через стеклянную призму выходящий поток лучей не является белым, а состоит из непрерывного спектра цветов, от фиолетового цвета на одном конце до красного на другом. Диапазон цветов, которые мы воспринимаем как видимый свет, составляет очень малую часть спектра электромагнитного излучения. На одном конце этого спектра находятся

радиоволны, длина которых в миллиарды раз превышает длины волн видимого света, а на другом конце - гамма-лучи, длина которых в миллионы раз меньше длины световых волн.

Электромагнитные волны можно трактовать как распространяющиеся синусоидальные колебания с длиной волны λ , а можно - как поток частиц с нулевой массой, движущихся со скоростью света. Каждая такая частица не имеет массы, но обладает определённой энергией и называется квантом излучения, или фотоном. Энергия фотона пропорциональна частоте излучения, т.е. электромагнитные колебания более высокой частоты (т.е. с более короткой длиной волны) обладают большей энергией фотона. Таким образом, у радиоволн энергия фотона мала, у микроволн энергия больше, у инфракрасного излучения еще больше, далее энергия фотона последовательно возрастает для диапазонов видимого спектра, ультрафиолетового излучения, рентгеновских лучей, и, наконец, гамма-лучей, обладающих самой большой энергией.

Свет является особым видом электромагнитного излучения, которое воспринимается человеческим глазом. На слайде приведён видимый цветовой спектр в "растянутом" виде. Видимый диапазон электромагнитного спектра охватывает длину волны приблизительно от 400 до 700 нм. Для удобства цветовой спектр часто делят на шесть широких цветовых диапазонов: фиолетовый, синий, зелёный, жёлтый, оранжевый и красный. Это деление достаточно условно - иногда говорят о пяти цветах, объединяя красно-оранжевый или сине-фиолетовый участки спектра, а иногда о семи цветах, выделяя голубой участок спектра между синим и зелёным. Ни один цветовой диапазон не имеет ярко выраженных границ, вместо этого каждый цвет плавно переходит в другой.

Итак, каждый источник света может быть однозначно определён его спектром - количеством излучаемой энергии для каждой длины волны.

Отражаемый свет Цвет, воспринимаемый человеком, как цвет объекта, определяется характером света, отражённого от этого объекта. Предмет, который отражает свет приблизительно одинаково во всём видимом диапазоне, представляется наблюдателю белым. Предмет, который отражает свет только в каком-то ограниченном диапазоне волн, будет восприниматься того цвета, который соответствует отражаемому диапазону волн. Например, красный предмет в основном отражает свет с длинами волн 625-740 нм, поглощая большинство энергии в других интервалах длин волн.

Свет, лишенный цветовой окраски, называется ахроматическим или монохроматическим. Единственным параметром такого освещения является его интенсивность, или яркость. Для описания монохроматической яркости также используется термин уровень серого поскольку яркость изменяется от черного до белого, с промежуточными оттенками серого.

Восприятие цвета: колбочки Рецепторами глаза, отвечающими за восприятие цветов человеком, являются колбочки. Экспериментально было

установлено, что все 6-7 миллионов колбочек в нашем глазу могут быть разделены на три основных группы по их восприимчивости к спектральному составу света. Эти три группы приблизительно соответствуют чувствительности к коротким, средним и длинным волнам, или синему, зелёному и красному цветам. Примерно 65% всех колбочек воспринимают красный цвет, 33% - зелёный и только около 2% - синий. На слайде показаны экспериментальные кривые спектральной чувствительности колбочек каждой из трёх групп для среднего нормального глаза.

Первичные и вторичные цвета Вследствие таких спектральных характеристик человеческий глаз воспринимает цвета как различные сочетания так называемых первичных цветов: красного, зелёного и синего. В 1931 г. Международная комиссия по освещению (МКО) разработала стандартный набор монохроматических первичных основных цветов, зафиксировав длины волн для стандартных синего, зелёного и красного цветов. Но важно понимать, что наличие стандартного набора монохроматических первичных основных цветов не означает, что все цвета спектра могут быть получены на основе этих фиксированных RGB цветов. Использование термина «основные» часто приводит к тому заблуждению, что все видимые цвета могут быть воспроизведены при смешении основных первичных цветов в различных пропорциях. Это не верно, за исключением того случая, когда длина волны основных цветов также может изменяться. Но в этом случае уже нельзя говорить о трёх стандартных первичных основных цветах.

Первичные основные цвета могут складываться, что даёт вторичные основные цвета: пурпурный (красный плюс синий), голубой (зелёный плюс синий) и жёлтый (красный плюс зелёный). Смешение трёх первичных основных цветов, или вторичного основного цвета и противоположного ему первичного, в правильных пропорциях даёт белый цвет. Результат такого смешения показан на слайде, где также показаны три первичных основных цвета и их сочетания, дающие вторичные основные цвета.

Также важно различать первичные основные цвета световых источников и первичные основные цвета красителей. Для красителей первичный основной цвет определяется как цвет, который поглощает один первичный основной цвет светового источника и отражает два оставшихся. Поэтому для красителей первичными основными цветами являются пурпурный, голубой и жёлтый, а вторичными — красный, зелёный и синий. Правильная комбинация трёх первичных основных цветов красителей или вторичного основного цвета и противоположного ему первичного даёт чёрный цвет.

Основные цвета световых источников ещё называют аддитивными, а красителей - субтрактивными.

1.6 Цифровое представление изображений

Как видит изображение компьютер Для формирования цифрового изображения необходимо преобразовать непрерывный аналоговый сигнал,

поступающий с сенсоров, в цифровую форму путём дискретизации и квантования. Давайте сначала рассмотрим монохромное изображение. Исходное аналоговое изображение непрерывно по координатам x и y , а также по уровню яркости. Чтобы преобразовать эту функцию в цифровую форму, необходимо ее представить отсчётами по обеим координатам и по уровням яркости. Представление координат в виде конечного множества отсчётов называется дискретизацией, а представление уровня яркости в значениями из конечного множества - квантованием.

В результате операций дискретизации и квантования возникает матрица действительных чисел. Такой формат представления изображения в виде матрицы называется растровым. Каждый элемент матрицы соответствует одному пикселю изображения. Значение элемента соответствует уровню яркости, или интенсивности, соответствующего пикселя. Значения интенсивности часто нормированы в диапазоне от 0 до 1. При такой нормировке белый пиксель будет иметь значение 1, чёрный - значение 0.

Параметры растрового изображения-1 Два основных параметра растрового изображения - это его пространственное и яркостное разрешение. Дискретизация является главным фактором, определяющим пространственное разрешение изображения. По существу, пространственное разрешение — это размер мельчайших различимых деталей на изображении. Пространственное разрешение (resolution) измеряется в числе пикселей на единицу пространственного измерения. Стандартная единица измерения - ppi, pixels per inch.

Размерность изображения тесно связана с его разрешением. Чем выше разрешение изображения, тем большего размера можно напечатать или отобразить на экране изображение без видимой потери качества.

Если нет необходимости измерять реальное пространственное разрешение и достигаемую степень детализации исходной сцены, то изображение, имеющее размеры $M \times N$ пикселей, часто называют изображением с пространственным разрешением $M \times N$ пикселей.

Уровни яркости Яркостным (или полутоновым) разрешением, называется мельчайшее различимое изменение яркости. Часто яркостное разрешение задают в числе уровней (или градаций). В виду особенностей цифровой техники, число уровней обычно выбирается равным степени 2. Наиболее частым решением является выбор 8-битного представления, что соответствует 256 уровням яркости. В некоторых приложениях используется 16 бит, если необходимо иметь более точное представление полутонов. Иногда можно встретить системы, в которых квантование уровней яркости изображения проводится с 10 или 12 битами точности, но это скорее исключение, чем правило.

На слайде приведены примеры изображений с разными уровнями яркости.

Параметры растрового изображения-2

Светлота, цветовой тон и насыщенность Часто для различения цветов используются такие параметры, как светлота, цветовой тон и насыщенность. Светлота является субъективной характеристикой. Она отражает уровень зрительного ощущения интенсивности света.

Цветовой тон характеризует доминирующий цвет, воспринимаемый наблюдателем. Когда мы называем некоторый объект красным, оранжевым или жёлтым, мы обозначаем его цветовой тон.

Насыщенность цвета связана с его относительной белизной. Спектрально чистые (мономатические) цвета являются полностью насыщенными. Такие цвета как розовый (смесь красного и белого) или бледно-лиловый (смесь фиолетового и белого) менее насыщены. Причём величина насыщенности цвета обратно пропорциональна количеству белого цвета в смеси.

Цветовой тон и насыщенность вместе называют цветностью. Каждый цвет может быть охарактеризован своей цветностью и светлотой.

Координаты цветности Величины красного, зелёного и синего, необходимые для получения некоторого конкретного цвета, называются координатами цвета и обозначаются соответственно X , Y и Z . Часто при описании цвета светлота не представляет интереса, и в таком случае цветовой тон и насыщенность можно выразить в координатах цветности x , y и z , которые определяются как

$$x = \frac{X}{X + Y + Z}, y = \frac{Y}{X + Y + Z}, z = \frac{Z}{X + Y + Z} \quad (1)$$

Из данных выражений видно, что $x + y + z = 1$.

Диаграмма цветностей Другой способ задавать цвета основан на использовании диаграммы цветностей, созданной Международной Комиссией по Освещению (МКО) в 1931г на основе серии экспериментов по оценке восприятия цвета человеком. Вся совокупность цветов может быть представлена как функция x (красной) и y (зелёной) координат цветности. Для любых значений координат x и y соответствующее значение синей координаты цветности z может быть получено из выражения $z = 1 - (x + y)$.

Вдоль границы диаграммы цветностей, имеющей форму языка, расположены различные цвета спектра — от фиолетового с длиной волны 380 нм до красного с длиной волны 780 нм. Любая точка, расположенная не на границе, а внутри диаграммы, представляет некоторую смесь цветов. Точка равной энергии, соответствующая координатам $(x, y) = (1/3, 1/3)$ соответствует равным долям трёх первичных основных цветов; она представляет опорный белый цвет стандарта МКО. Любая точка, расположенная на границе диаграммы цветностей, имеет максимальную цветовую насыщенность. По мере того как точка смещается от границы к точке равной энергии, соответствующий ей цвет содержит в своём составе все большую долю белого

и становится все менее насыщенным. Цветовая насыщенность точки равной энергии равна нулю.

Диаграмма цветности полезна при рассмотрении процедуры смешения цветов, поскольку отрезок, соединяющий любые две точки диаграммы, определяет всевозможные различные цвета, которые могут быть получены при смешении двух данных цветов. Аналогично, отрезок, проведённый от точки равной энергии к любой точке границы диаграммы, определяет все оттенки выбранного цвета.

Цветовой охват Такая процедура обобщается и на случай смешения трёх цветов. Для того чтобы определить диапазон цветов, которые могут быть получены комбинацией трёх любых заданных цветов, нужно просто соединить между собой отрезками соответствующие точки на диаграмме цветности. В результате получится треугольник, и все цвета, соответствующие точкам внутри этого треугольника, могут быть получены как различные комбинации трёх первоначальных цветов.

Заметим, что никакой треугольник с вершинами в трёх точках с неизменными цветами не может включать весь диапазон цветов диаграммы цветностей. Это означает, что нельзя получить все цвета с помощью трёх фиксированных первичных основных цветов.

Треугольник, отмеченный правом рисунке, представляет типичный для RGB мониторов диапазон воспроизводимых цветов, который называют цветовым охватом. Область сложной формы внутри этого треугольника представляет типичный цветовой охват печатающих устройств. Граница области охвата для печатающих устройств имеет сложную форму, потому что в процессе цветной печати одновременно используются аддитивные и субтрактивные процедуры смешения цветов. Управлять таким процессом намного труднее, чем процессом воспроизведения цветов на экране монитора, основанном на смешении трёх хорошо контролируемых первичных основных цветов.

1.7 Модели цвета

Для того, чтобы сделать возможным описание цветов некоторым стандартным способом, введём понятие цветового пространства. Цветовое пространство также иногда называют цветовой моделью или системой цветов. По существу, цветовое пространство определяет некоторую систему координат и подпространство внутри этой системы, в котором каждый цвет представляется единственной точкой.

Большинство современных цветовых пространств ориентированы либо на устройства цветовоспроизведения (например, цветные мониторы или принтеры), либо на определённые прикладные задачи обработки изображений. Аппаратно-ориентированными цветовыми моделями, наиболее часто используемыми на практике, являются модель RGB для цветных мониторов и широкого класса цветных видеокамер, модели CMY и CMYK для цветных принтеров, и модели, разделяющие цветовую и яркостную информацию

на изображении, такие как Hue-Saturation-Intensity (HSI), Hue-Saturation-Value (HSV), Hue-Saturation-Lightness (HSL), Hue-Saturation-Brightness (HSB). Такие модели лучше соответствуют цветовосприятию человека и дают возможность применять многие из полутоновых методов обработки изображений и для цветных изображений.

Ещё одно популярное цветовое пространство, используемое в задачах обработки изображений - CIE LAB, в котором изменение цвета более линейно с точки зрения человеческого восприятия.

Далее рассмотрим более подробно некоторые из этих пространств.

RGB Наверное, самым распространенным цветовым пространством является пространство RGB. В данной модели каждый цвет представляется красным, зеленым и синим первичными основными цветами. В основе модели лежит декартова система координат. Цветовое пространство представляет собой куб. Точки, отвечающие красному, зелёному и синему цветам, расположены в трех вершинах куба, лежащих на координатных осях. Голубой, пурпурный и желтый цвета расположены в трёх других вершинах куба. Чёрный цвет находится в начале координат, а белый — в наиболее удалённой от начала координат вершине. В рассматриваемой модели оттенки серого цвета (точки с равными RGB значениями) лежат на диагонали, соединяющей чёрную и белую вершины. Различные цвета в этой модели представляют собой точки на поверхности или внутри куба и определяются вектором, проведённым в данную точку из начала координат. Для удобства предполагается, что все значения цвета нормированы таким образом, чтобы куб был единичным кубом, т.е. все значения R, G и B лежат в диапазоне $[0,1]$.

Представляемые в цветовой модели RGB изображения состоят из трёх отдельных изображений-компонент, по одному для каждого первичного основного цвета. При воспроизведении RGB монитором эти три изображения смешиваются на люминесцирующем экране и образуют составное цветное изображение. Число битов, используемых для представления каждого пикселя в RGB пространстве, называется глубиной цвета. Рассмотрим RGB изображение, в котором каждая из компонент — красная, зеленая и синяя — является 8-битовой. В таком случае говорят, что каждый цветной RGB пиксель (т.е. триплет значений (R, G, B)) имеет глубину 24 бита (три цветовые плоскости умножить на число битов на каждую плоскость); для такого изображения часто используется термин полноцветное изображение. Суммарное число всевозможных цветов в 24-битовом RGB изображении составляет $(2^8)^3 = 16777216$.

Цветовая система Манселла Цветовая (или колориметрическая) система Манселла — цветовое пространство, разработанное профессором Альбертом Манселлом в начале XX века. Цвет в нем описывается с помощью трёх чисел — цветового тона, значения (светлоты) и хромы (насыщенности). Попытки создать цветовое пространство, цвет в котором описывался

бы тремя координатами, были и до Манселла, но он первым решил разделить цвет на независимые значения тона, светлоты и насыщенности. Его система, особенно её поздние редакции, основывалась на тщательных экспериментах по изучению цветового восприятия человека, то есть под неё была подведена научная основа. Благодаря этому, цветовая система Манселла пережила многие системы того времени, и, хотя в большинстве приложений её заменили более современные системы, такие как (CIE L^*a^*b), она все ещё применяется в некоторых областях. Например, в стандартах ANSI для определения цвета кожи и волос человека, в судебной медицине, в геологии для сравнения цвета почвы и в пивоварении для определения цвета пива.

Цветовая система Манселла включает три координаты. Цветовое тело можно представить как цилиндр в трёхмерном пространстве. Цветовой тон измеряется в градусах по горизонтальной окружности, хрома (насыщенность) измеряется радиально от нейтральной оси цилиндра к более насыщенным краям, значение (светлота) измеряется вертикально по оси цилиндра от 0 (черный) до 10 (белый). Расположения цветов определялось экспериментально изучением цветового ощущения испытуемых. Цвета Манселл пытался расположить визуально одинаково, что привело к образованию цветового тела неправильной формы.

Каждый горизонтальный круг в системе Манселла разделен на пять основных тонов: красный (Red), желтый (Yellow), зеленый (Green), голубой (Blue) и фиолетовый (Purple). Между ними располагаются пять переходных тонов. В нижней части слайда показаны цветовые тона по Манселлу для фиксированных значений насыщенности и светлоты. Два цвета одинаковой светлоты и насыщенности на противоположных сторонах круга смешиваются в нейтральный серый того же значения.

HSI Другими цветовыми пространствами, используемыми в обработке изображений и схожими с системой Манселла, является семейство пространств HSI/HSL/HSV/HSB. Цветовое пространство RGB идеально приспособлено для аппаратной реализации и удачно согласовано со зрительной системой человека в том смысле, что человеческий глаз восприимчив к красному, зеленому и синему — первичным основным цветам. Но пространство RGB плохо приспособлено для описания цветов так, как это свойственно человеку. Например, описывая цвет автомобиля, мы не говорим о процентном содержании в нем красного, зелёного и синего. И рассматривая цветное изображение мы не думаем о том, что оно составлено из трёх отдельных изображений — по одному для каждого первичного основного цвета. Глядя на окрашенный объект, человек описывает его с помощью цветового тона, насыщенности и светлоты.

Пространство HSI и другие схожие пространства позволяют отделить информацию о цветности (тон и насыщенность) от информации о яркости (интенсивность). Поскольку в основе модели лежит естественное и интуитивно близкое человеку описание цвета, пространство HSI часто использовалось для построения алгоритмов обработки изображений.

HSI/HSL/HSV/HSB Все пространства этого семейства могут быть представлены, как геометрическое преобразование RGB куба. Если поставить RGB куб на угол, соответствующий черному цвету с осью-диагональю от черного к белому, то тогда насыщенность можно измерять, как расстояние от этой оси - чем дальше от оси, тем насыщеннее цвет. А цветовой тон можно измерять как угол вокруг этой оси, начиная с угла в 0° , соответствующего красному цвету. Основное отличие всех пространств этого семейства лежит в определении яркостной составляющей. Интенсивность I в пространстве HSI определяется как среднее арифметическое R , G и B координат в исходном пространстве RGB. Светлота L (lightness) в пространстве HSL определяется, как среднее арифметическое максимума и минимума значений R , G , B . Значение V (value) определяется как максимум R , G , B .

CIE Lab Цветовое пространство XYZ (или, что тоже самое, диаграмма цветностей МКО) вмещает в себя все воспринимаемые человеком цвета. Но данное пространство нелинейно, или, иными словами, зрительно неоднородно. Это значит, что два равноудалённых цвета от некоторой точки в данном пространстве необязательно будут восприниматься зрительно как равноудалённые от цвета, соответствующего этой некоторой точке. Было предпринято несколько попыток устранить нелинейность XYZ. В итоге в 1976 году была разработана модель CIE $L^*a^*b^*$, которая является сейчас международным стандартом.

В цветовом пространстве Lab значение светлоты отделено от значения хроматической составляющей цвета (тон, насыщенность). Светлота задана координатой L (изменяется от 0 до 100, то есть от самого тёмного до самого светлого), хроматическая составляющая — двумя декартовыми координатами a и b . Первая обозначает положение цвета в диапазоне от зелёного до красного, вторая — от синего до жёлтого.

В отличие от цветовых пространств RGB или CMYK, которые являются, по сути, набором аппаратных данных для воспроизведения цвета на бумаге или на экране монитора, и в которых цвет может зависеть от типа печатной машины или марки красок, или производителя монитора и его настроек, CIE Lab однозначно определяет цвет. Поэтому данное пространство нашло широкое применение в программном обеспечении для обработки изображений в качестве промежуточного цветового пространства, через которое происходит конвертирование данных между другими цветовыми пространствами. Помимо этого, особые свойства Lab сделали редактирование в этом пространстве мощным инструментом цветокоррекции. Благодаря характеру определения цвета в Lab появляется возможность отдельно воздействовать на яркость, контраст изображения и на его цвет. Во многих случаях это позволяет ускорить обработку изображений. Lab предоставляет возможность избирательного воздействия на отдельные цвета в изображении, усиления цветового контраста.

HCL И ещё одним цветовым пространством, которое мы кратко упомянем, является пространство HCL. Это пространство было создано, чтобы совместить в себе свойства пространств, полученных при помощи цилиндрических преобразований RGB (пространств семейства HSI) и свойства линейности цвета пространства CIE Lab.

Представление цифровых изображений Как же компьютер представляет и хранит цветное изображение? Ранее мы говорили, что монохромное черно-белое растровое изображение представляет собой матрицу, в которой значение каждого элемента соответствует уровню яркости, или интенсивности, соответствующего пикселя.

Представление цветного изображения Для хранения информации о цвете изображения нам необходимо хранить информацию о цвете каждого пикселя. Поскольку большинство цветовых пространств позволяет задать цвет при помощи трёх координат, цветное растровое изображение будет хранить значения трёх цветовых координат для каждого пикселя. Так, при представлении изображения в пространстве RGB, для каждого пикселя будет храниться значения красной, зелёной и синей составляющих данного цвета. Т.е. цветное изображение, в отличие от монохромного, может быть представлено трехмерным тензором, в котором одна из размерностей соответствует координатам цвета (или цветовым каналам).