СИСТЕМЫ КОМПЬЮТЕРНОГО ЗРЕНИЯ: СОВРЕМЕННЫЕ ЗАДАЧИ И МЕТОДЫ

АЛЕКСЕЙ ПОТАПОВ
pas.aicv@gmail.com

Область компьютерного зрения является действительно инновационно привлекательной. Интерес к ней возник на заре попыток создания искусственного интеллекта. В настоящее время количество новых решений и актуальных приложений для компьютерного зрения продолжает расти.

ТРУДНАЯ ПРОБЛЕМА ЗРЕНИЯ

Интерес к компьютерному зрению возник одним из первых в области искусственного интеллекта наряду с такими задачами, как автоматическое доказательство теорем и интеллектуальные игры. Даже архитектура первой искусственной нейронной сети — перцептрона — была предложена Фрэнком Розенблаттом, исходя из аналогии с сетчаткой глаза, а ее исследование проводилось на примере задачи распознавания изображений символов.

Значимость проблемы зрения никогда не вызывала сомнения, но одновременно ее сложность существенно недооценивалась. К примеру, легендарным по своей показательности стал случай, когда в 1966 г. один из основоположников области искусственного интеллекта, Марвин Минский, даже не сам собрался решить проблему искусственного зрения, а поручил это сделать одному студенту за ближайшее лето [1]. При этом на создание программы, играющей на уровне гроссмейстера в шахматы, отводилось значительно большее время. Однако сейчас очевидно, что создать программу, обыгрывающую человека в шахматы, проще, чем создать адаптивную систему управления с подсистемой компьютерного зрения, которая бы смогла просто переставлять шахматные фигуры на произвольной реальной доске.

Прогресс в области компьютерного зрения определяется двумя факторами: развитие теории, методов, и развитие аппаратного обеспечения. Долгое время теория и академические исследования опережали возможности практического использования систем компьютерного зрения. Условно можно выделить ряд этапов развития теории.

- К 1970-м годам сформировался основной понятийный аппарат в области обработки изображений, являющийся основой для исследования проблем зрения. Также были выделены основные задачи, специфические для машинного зрения, связанные с оценкой физических параметров сцены (дальности, скоростей движения, отражательной способности поверхностей и т. д.) по изображениям, хотя ряд этих задач все еще рассматривался в весьма упрощенной постановке для «мира игрушечных кубиков».
- К 80-м сформировалась теория уровней представления изображений в методах их анализа. Своего рода отметкой окончания этого этапа служит книга Дэвида Марра «Зрение. Информационный подход к изучению представления и обработки зрительных образов».
- К 90-м оказывается сформированным систематическое пред-

ставление о подходах к решению основных, уже ставших классическими, задач машинного зрения.

- С середины 90-х происходит переход к созданию и исследованию крупномасштабных систем компьютерного зрения, предназначенных для работы в различных естественных условиях.
- Текущий этап наиболее интересен развитием методов автоматического построения представлений изображений в системах распознавания изображений и компьютерного зрения на основе принципов машинного обучения.

В то же время прикладные применения ограничивались вычислительными ресурсами. Ведь чтобы выполнить даже простейшую обработку изображения, нужно хотя бы один раз просмотреть все его пиксели (и обычно не один раз). Для этого нужно выполнять как минимум сотни тысяч операций в секунду, что долгое время было невозможно и требовало упрощений.

К примеру, для автоматического распознавания деталей в промышленности могла использоваться черная лента конвейера, устраняющая необходимость отделения объекта от фона, или сканирование движущегося объекта линейкой фотодиодов со специальной подсветкой, что уже на уровне формирования сигнала обеспечивало выделение инвариантных признаков для распознавания без применения каких-либо сложных методов анализа информации. В оптико-электронных системах сопровождения и распознавания целей использовались физические трафареты, позволяющие «аппаратно» выполнять согласованную фильтрацию. Некоторые из этих решений являлись гениальными с инженерной точки зрения, но были применимы только в задачах с низкой априорной неопределенностью, и поэтому обладали, в частности, плохой переносимостью на новые задачи.

Не удивительно, что на 1970-е годы пришелся пик интереса и к оптическим вычислениям в обработке изображений. Они позволяли реализовать небольшой набор методов (преимущественно корреляционных) с ограниченными свойствами инвариантности, но весьма эффективным образом.

Постепенно, благодаря росту производительности процессоров (а также развитию цифровых видеокамер), ситуация изменилась. Преодоление определенного порога производительности, необходимого для осуществления полезной обработки изображений за разумное время, открыло путь для целой лавины приложений компьютерного зрения. Следует, однако, сразу подчеркнуть, что этот переход не был мгновенным и продолжается до сих пор.

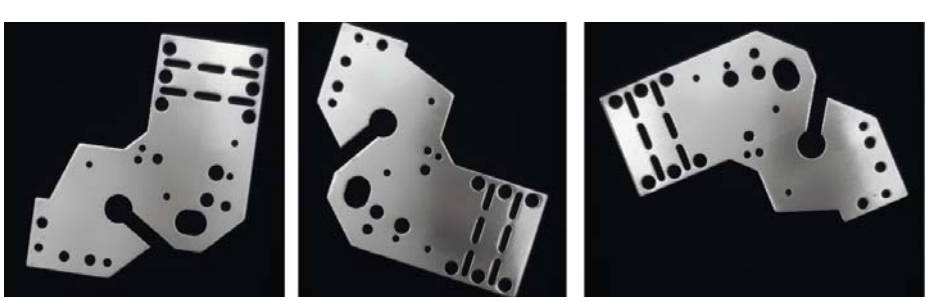
В первую очередь, общеприменимые алгоритмы обработки изображений стали доступны для спецпроцессоров — цифровых сигнальных процессоров (ЦСП) и программируемых логических интегральных схем (ПЛИС), нередко совместно использовавшихся и находящих широкое применение до сих пор в бортовых и промышленных системах.

Однако действительно массовое применение методы компьютерного зрения получили лишь менее десяти лет назад, с достижением соответствующего уровня производительности процессоров у персональных и мобильных компьютеров. Таким образом, в плане практического применения системы компьютерного зрения прошли ряд этапов: этап индивидуального решения (как в части аппаратного обеспечения, так и алгоритмов) конкретных задач; этап применения в профессиональных областях (в особенности в промышленности и оборонной сфере) с использованием спецпроцессоров, специализированные системы формирования изображений и алгоритмы, предназначенные для работы в условиях низкой априорной неопределенности, однако эти решения допускали масштабирование; и этап массового применения.

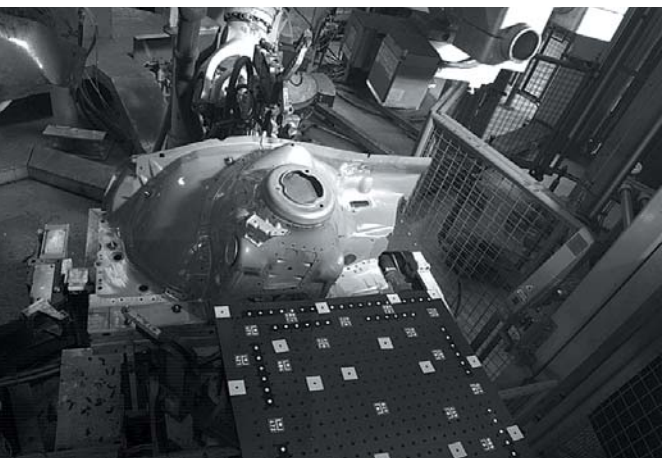
Как видно, система машинного зрения включает следующие основные компоненты:

- подсистему формирования изображений (которая сама может включать разные компоненты, например объектив и ПЗС- или КМОП-матрицу);
- вычислитель;
- алгоритмы анализа изображений, которые могут реализовываться программно на процессорах общего назначения, аппаратно в структуре вычислителя и даже аппаратно в рамках подсистемы формирования изображений.

Наиболее массового применения достигают системы машинного зрения, использующие стандартные камеры и компьютеры в качестве первых двух компонент (именно к таким системам больше подходит термин «компьютерное зрение», хотя четкого разделения понятий машинного и компьютерного зрения нет). Однако, естественно, прочие системы машинного зрения обладают не меньшей значимостью. Именно выбор «нестандартных» способов формирования изображений (включая использование иных, помимо видимого, спектральных диапазонов, когерентного излучения, структурированной подсветки, гиперспектральных приборов, времяпролетных, всенаправленных и быстродействующих камер, телескопов и микроскопов и т. д.) существенно расширяет возможности систем машинного зрения. В то время как по возможностям алгоритмического обеспечения системы машинного зрения существенно уступают зрению человека, по возможностям получения информации о наблюдаемых объектах они существенно превосходят его. Однако вопросы формирования изображений составляют самостоятельную область, а методы работы

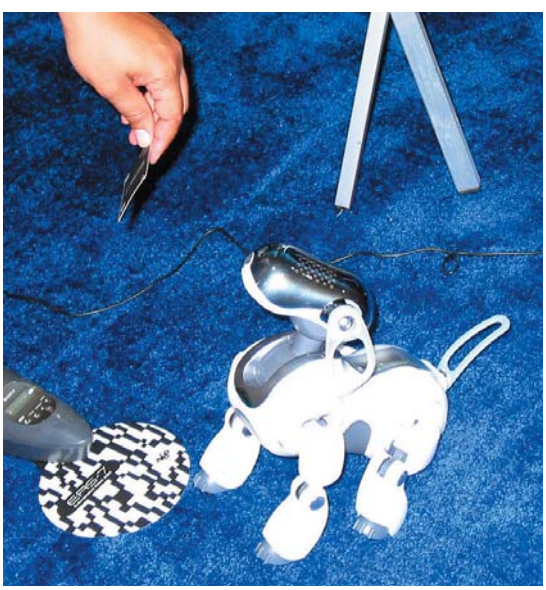


▼ Максимальная изменчивость внешнего вида детали на ленте конвейера

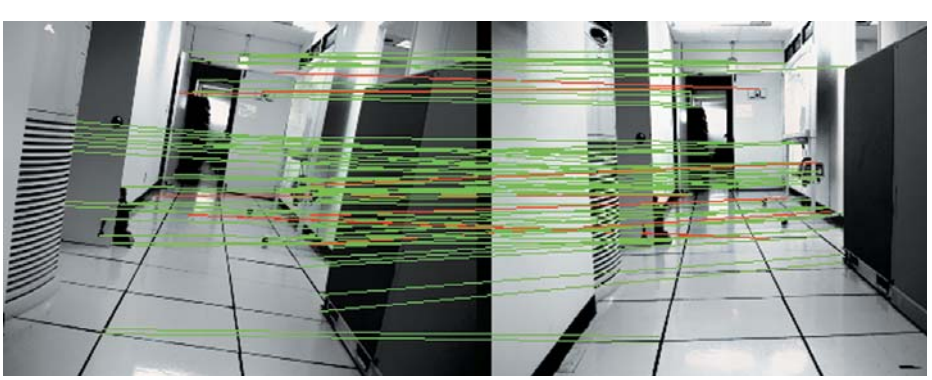


▲ Использование доски с фотометрическими метками для определения внешней ориентации камер

▼ Робот AIBO распознает знак на карточке как команду для выполнения



▼ Сопоставление изображений внутри помещения в целях узнавания местоположения



с изображениями, полученными с использованием разных сенсоров, столь разнообразны, что их обзор выходит за рамки данной статьи. В этой связи мы ограничимся обзором систем компьютерного зрения, использующих обычные камеры.

ПРИМЕНЕНИЕ В РОБОТОТЕХНИКЕ

Робототехника является традиционной областью применения машинного зрения. Однако основная доля парка роботов долгое время приходилась на промышленность, где очувствление роботов не было лишним, но благодаря хорошо контролируемым условиям (низкой недетерминированности среды) возможными оказывались узкоспециализированные решения, в том числе и для задач машинного зрения. Кроме того, промышленные приложения допускали использование дорогостоящего оборудования, включающего оптические и вычислительные системы.

В этой связи показательно (хотя и не связано только с системами компьютерного зрения) то, что доля парка роботов, приходящаяся на промышленных роботов, стала менее 50% лишь в начале 2000-х годов [2]. Стала развиваться робототехника, предназначенная для массового потребителя. Для бытовых роботов, в отличие от промышленных, критичной является стоимость, а также время автономной работы, что подразумевает использование мобильных и встраиваемых процессорных систем. При этом такие роботы должны функционировать в недетерминированных средах. К примеру, в промышленности долгое время (да и по сей день) использовались фотограмметрические метки, наклеиваемые на объекты наблюдения или калибровочные доски, — для решения задач определения внутренних параметров и внешней ориентации камер. Естественно, необходимость наклеивать пользователю такие метки на предметы интерьера существенно ухудшила бы потребительские качества бытовых роботов. Не удивительно, что рынок бытовых роботов ждал для начала своего бурного развития достижения определенного уровня технологий, что произошло в конце 90-х.

Точкой отсчета этого события может служить выпуск первой версии робота AIBO (Sony), который, несмотря на сравнительно высокую цену ($2500), пользовался большим спросом. Первая партия этих роботов в количестве 5000 экземпляров была раскуплена в Интернете за 20 мин., вторая партия (также в 1999 г.) — за 17 с, и далее темп продаж составлял порядка 20 000 экземпляров в год.

Также в конце 90-х появились в массовом производстве устройства, которые можно было бы назвать бытовыми роботами в полном смысле этого слова. Наиболее типичными автономными бытовыми роботами являются роботы-пылесосы. Первой моделью, выпущенной в 2002 г. фирмой iRobot, стала Roomba. Затем появились роботы-пылесосы, выпущенные фирмами LG Electronics, Samsung и др. К 2008 г. суммарные объемы продаж роботов-пылесосов в мире составили более полумиллиона экземпляров в год.

Показательно то, что первые роботы-пылесосы, оснащенные системами компьютерного зрения, появились лишь в 2006 г. К этому моменту использование мобильных процессоров типа семейства ARM с частотой 200 МГц позволяло добиться сопоставления изображений трехмерных сцен внутри помещений на основе инвариантных дескрипторов ключевых точек в целях сенсорной локализации робота с частотой порядка 5 кадров/с. Использование зрения для определения роботом своего местоположения стало экономически оправданным, хотя еще недавно для этих целей производители предпочитали использовать сонары.

Дальнейшее повышение производительности мобильных процессоров позволяет ставить новые задачи для систем компьютерного зрения в бытовых роботах, число продаж которых по всему миру исчисляется уже миллионами экземпляров в год [3]. Помимо задач навигации, от роботов, предназначенных для персонального использования, может потребоваться решение задач распознавания людей и их эмоций по лицам, распознавание жестов, предметов обстановки, включая столовые приборы и посуду, одежду, домашних животных и т. д., в зависимости от типа задачи, решаемой роботом. Многие из этих задач далеки от полного решения и являются перспективными с инновационной точки зрения.

Таким образом, современная робототехника требует решения широкого круга задач компьютерного зрения, включающего, в частности:

- набор задач, связанных с ориентацией во внешнем пространстве (например, задачу одновременной локализации и картографирования — Simultaneous Localization and Mapping, SLAM), определением расстояний до объектов и т. д;
- задачи по распознаванию различных объектов и интерпретации сцен в целом;
- задачи по обнаружению людей, распознаванию их лиц и анализу эмоций.

СИСТЕМЫ ПОМОЩИ ВОДИТЕЛЮ

Помимо бытовых роботов, методы компьютерного зрения нашли широкое применение в системах помощи водителю. Работы по детектированию разметки, препятствий на дороге, распознаванию знаков и т. д. активно велись и в 90-х годах. Однако достаточного уровня (как по точности и надежности самих методов, так и по производительности процессоров, способных в масштабе реального времени выполнять соответствующие методы) они достигли преимущественно в последнем десятилетии.

Одним из показательных примеров являются методы стереозрения, используемые для обнаружения препятствий на дороге. Эти методы могут быть весьма критичны к надежности, точности и производительности. В частности, в целях обнаружения пешеходов может требоваться построение плотной карты дальности в масштабе, близком к реальному времени. Эти методы могут требовать сотен операций на пиксель и точности, достигаемой при размерах изображений не менее мегапиксела, то есть при сотнях миллионов операций на кадр (нескольких миллиардов и более операций в секунду).

Стоит отметить, что общий прогресс в области компьютерного зрения отнюдь не связан только с развитием аппаратного обеспечения. Последнее лишь открывает возможности для применения вычислительно затратных методов обработки изображений, но сами эти методы также нуждаются в разработке. За последние 10–15 лет были доведены до эффективного практического использования методы сопоставления изображений трехмерных сцен [4, 5], методы восстановления плотных карт дальности на основе стереозрения [6], методы обнаружения и распознавания лиц [7] и т. д. Общие принципы решения соответствующих задач данными методами не изменились, но они обогатились рядом нетривиальных технических деталей и математических приемов, сделавших эти методы успешными.

Возвращаясь к системам помощи водителю, нельзя не упомянуть про современные методы детектирования пешеходов, в частности, на основе гистограмм ориентированных градиентов [8]. Современные методы машинного обучения, о которых еще будет сказано позднее, впервые позволили компьютеру лучше человека решать такую достаточно общую зрительную задачу, как распознавание дорожных знаков [9], но не благодаря использованию специальных средств формирования изображений, а благодаря алгоритмам распознавания, получавшим на вход в точности ту же информацию, что и человек.

Одним из существенных технических достижений стал беспилотный автомобиль Google, который, однако, использует богатый набор сенсоров помимо видеокамеры, а также не работает на незнакомых (заранее не отснятых) дорогах и при плохих погодных условиях.

◀ Детектирование ключевых точек на лице человека для распознавания эмоций

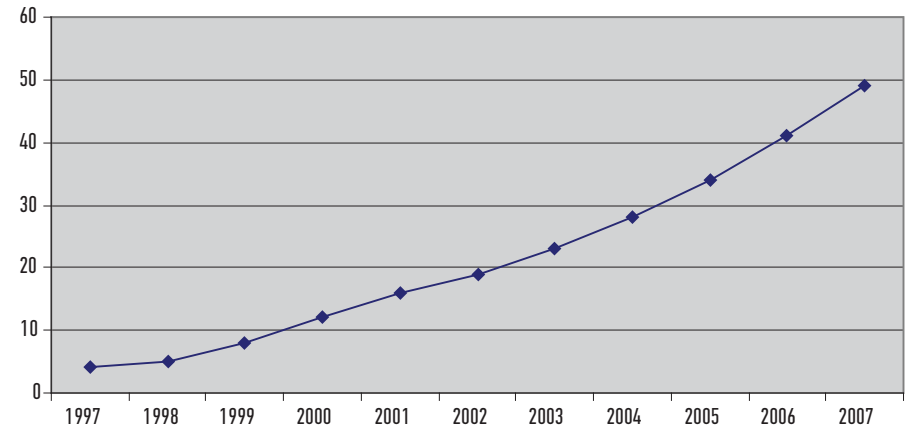
Таким образом, для систем помощи водителю требуется решение разных задач компьютерного зрения, включая:

- стереозрение;
- обнаружение препятствий на дорогах;
- распознавание дорожных знаков, разметки, пешеходов и автомобилей;
- задачи, также требующие упоминания, связанные с контролем состояния водителя.

МОБИЛЬНЫЕ ПРИЛОЖЕНИЯ

Еще более массовыми по сравнению с бытовой робототехникой и системами помощи водителю являются задачи компьютерного зрения для персональных мобильных устройств, таких как смартфоны, планшеты и т. д. В частности, число мобильных телефонов

▼ Стереозрение для детектирования препятствий на дороге



▲ Количество владельцев мобильных телефонов на 100 жителей

неуклонно растет и уже практически превысило по численности население Земли. При этом основная доля телефонов выпускается сейчас с камерами. В 2009 г. количество таких телефонов превысило миллиард, что создает колоссальный по размерам рынок для систем обработки изображений и компьютерного зрения, который далек от насыщения, несмотря на многочисленные R&D-проекты, проводящиеся как самими фирмами — изготовителями мобильных устройств, так и большим числом стартапов.

Часть задач по обработке изображений для мобильных устройств с камерами совпадает с задачами для цифровых фотоаппаратов. Основное отличие заключается в качестве объективов и в условиях съемки. Для примера можно привести задачу синтеза изображений с расширенным динамическим диапазоном (HDRI) по нескольким снимкам, полученным с разной экспозицией. В случае мобильных устройств на изображениях присутствует больший шум, кадры формируются с большим интервалом времени, и смещение камеры в пространстве также больше, что усложняет задачу получения качественных HDRI-изображений, которую при этом приходится решать на процессоре мобильного телефона. В этой связи решение, казалось бы, идентичных задач для разных устройств может различаться, что делает эти решения до сих пор востребованными на рынке.

▼ Сопоставление изображения постера фильма, снятого камерой мобильного телефона, с эталоном

Больший интерес, однако, представляют новые приложения, которые ранее отсутствовали на рынке. Широкий класс таких приложений для персональных мобильных устройств связан с задачами дополненной реальности, которые могут быть весьма разнообразными. Сюда относятся игровые приложения (требующие согласованного отображения виртуальных объектов поверх изображения реальной сцены при перемещении камеры), а также различные развлекательные приложения в целом, туристические приложения (распознавание достопримечательностей с выводом информации о них), а также многие другие приложения, связанные с информационным поиском и распознаванием объектов: распознавание надписей на иностранных языках с отображением их перевода, распознавание визитных карточек с автоматическим занесением информации в телефонную книгу, а также распознавание лиц с извлечением информации из телефонной книги, распознавание постеров фильмов (с заменой изображения постера на трейлер фильма) и т. д.

Системы дополненной реальности могут создаваться в виде специализированных устройств типа Google Glass, что еще больше увеличивает инновационный потенциал методов компьютерного зрения.

Таким образом, класс задач компьютерного зрения, решения которых могут быть применены в мобильных приложениях, крайне широк. Большой набор приложений есть у методов сопоставления (отождествления сопряженных точек) изображений, в том числе с оценкой трехмерной структуры сцены и определением изменения ориентации камеры и методов распознавания объектов, а также анализа лиц людей. Однако может быть предложено неограниченно большое число мобильных приложений, для которых будет требоваться разработка специализированных методов компьютерного зрения. Приведем лишь два таких примера: запись на мобильный телефон с автоматической дешифрацией партии в некоторой настольной игре и реконструкция траектории движения клюшки для гольфа при нанесении удара.

ИНФОРМАЦИОННЫЙ ПОИСК И ОБУЧЕНИЕ

Многие задачи дополненной реальности тесно связаны с информационным поиском (так что некоторые системы, такие как Google Goggles, сложно отнести к какой-то конкретной области), который представляет существенный самостоятельный интерес.

Задачи поиска изображений по содержанию также разнообразны. Они включают сопоставление изображений при поиске изображений уникальных объектов, например архитектурных сооружений, скульптур, картин и т. д., обнаружение и распознавание на изображениях объектов классов разной степени общности (автомобилей, животных, мебели, лиц людей и т. д., а также их подклассов), категоризация сцен (город, лес, горы, побережье и т. д.). Эти задачи могут встречаться в различных приложениях — для сортировки изображений в домашних цифровых фотоальбомах, для поиска товаров по их изображениям в интернет-магазинах, для извле-

чения изображений в геоинформационных системах, для систем биометрической идентификации, для специализированного поиска изображений в социальных сетях (например, поиска лиц людей, привлекательных для пользователя) и т. д., вплоть до поиска изображений в Интернете.

Как уже достигнутый прогресс, так и перспективы его продолжения видны на примере конкурса Large Scale Visual Recognition Challenge [10], в котором количество распознаваемых классов увеличилось с 20 в 2010 г. до 200 в 2013-м.

Распознавание объектов стольких классов сейчас немыслимо без привлечения методов машинного обучения в область компьютерного зрения. Одно из крайне популярных направлений здесь — сети глубокого обучения, предназначенные для автоматического построения многоуровневых систем признаков, по которым происходит дальнейшее распознавание. Востребованность этого направления видна по фактам приобретения различных стартапов такими корпорациями, как Google и Facebook. Так, корпорацией Google в 2013 г. была куплена фирма DNNresearch, а в начале 2014 г. — стартап DeepMind. Причем за покупку последнего стартапа конкурировал и Facebook (который до этого нанял такого специалиста, как Ян Ле Кун, для руководства лабораторией, ведущей разработки в области глубокого обучения), а стоимость покупки составила $400 млн. Стоит отметить, что и упоминавшийся метод [8], выигравший в конкурсе по распознаванию дорожных знаков, также основан на сетях глубокого обучения.

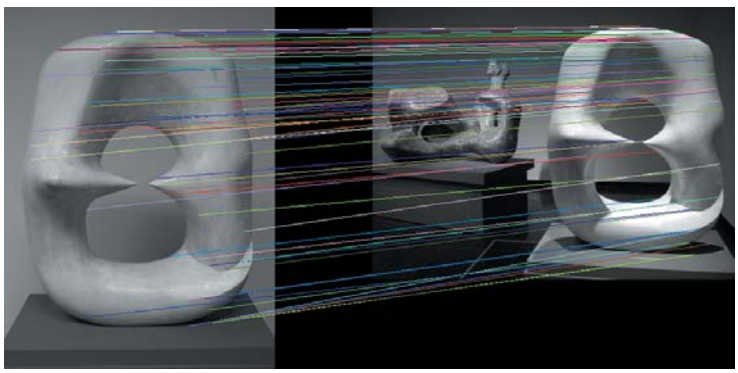
Методы глубокого обучения требуют огромных вычислительных ресурсов, и даже для обучения распознаванию ограниченного класса объектов могут требоваться несколько дней работы на вычислительном кластере. При этом в будущем могут быть разработаны еще более мощные, но требующие еще больших вычислительных ресурсов методы.

ЗАКЛЮЧЕНИЕ

Мы рассмотрели лишь наиболее распространенные приложения компьютерного зрения для массового пользователя. Однако существует и множество других, менее типичных приложений. К примеру, методы компьютерного зрения могут быть использованы в микроскопии, оптической когерентной томографии, цифровой голографии. Многочисленны приложения методов обработки и анализа изображений в различных профессиональных областях — биомедицине, космической отрасли, криминалистике и т. д.

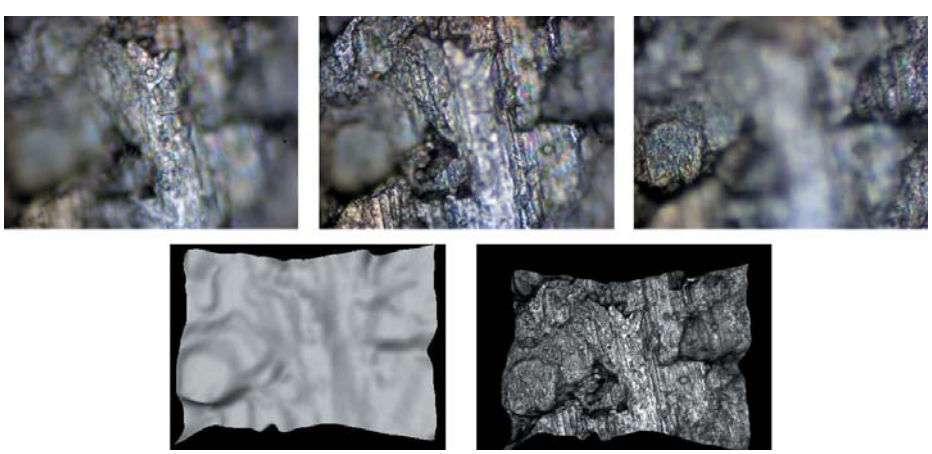
В настоящее время количество актуальных приложений компьютерного зрения продолжает расти. В частности, для решения становятся доступными задачи, связанные с анализом видеоданных. Активное развитие трехмерного телевидения расширяет заказ на системы компьютерного зрения, для создания которых не разработаны еще эффективные алгоритмы и требуются более существенные вычислительные мощности. Такой востребованной задачей является, в частности, задача конвертации видео 2D в 3D.

Неудивительно, что на переднем фронте систем компьютерного зрения продолжают активно использоваться специальные вычислительные средства. В частности, сейчас популярны графические процессоры общего назначения (GPGPU) и облачные вычисления. Однако соответствующие решения постепенно перетекают в сегмент персональных компьютеров с существенным расширением возможных приложений.



◀ Распознавание скульптур

▼ Распознавание объектов на сцене



▼ Восстановление 3D-профиля листа металла, наблюдаемого с помощью микроскопа, методом «глубина из фокусировки»

Таким образом, область компьютерного зрения является действительно инновационно привлекательной. В то же время она является весьма наукоемкой и требует проведения НИОКР, уровень сложности которых может превышать возможности обычных стартапов в области информационных технологий. Преимуществом при решении задач здесь могут обладать коллективы и лаборатории, функционирующие в университетах и научных организациях. Именно к таковым относится, в частности, международный научно-технический центр Вычислительной оптики, фотоники и визуализации изображений, организованный силами сотрудников кафедры компьютерной фотоники и видеоинформатики НИУ ИТМО с привлечением специалистов отдела обработки изображений ОАО «ГОИ им. С. И. Вавилова» и решающий, среди прочих, практически все упоминавшиеся выше задачи компьютерного зрения и обработки изображений. Научно-исследовательская и инновационная деятельность центра также тесно интегрирована с образовательным процессом. Ведь задачи компьютерного зрения являются увлекательными, но сложными, и чтобы будущие специалисты смогли закрепиться в данной области, им необходимо преодолеть данный порог сложности. Для этого необходимо приобрести как базовые навыки и знания, так и представление о современных методах и нерешенных проблемах, что реализуется с помощью читаемых на кафедре дисциплин. Однако помимо этого необходим личный опыт участия в решении реальных, а не сугубо учебных задач, позволяющих студентам ощутить себя на острие идущего научно-технического прогресса, чтобы осознать, что они спустя некоторое время вполне могут стать его движущей силой. ●

ЛИТЕРАТУРА

1. Bechtel W. The Cardinal Mercier Lectures at the Catholic University of Louvain: An Exemplar Neural Mechanism: The Brain's Visual Processing System. 2003.
2. Юревич Е. И. Основы робототехники. 2-е изд. СПб: БХВ-Петербург. 2007.
3. Executive Summary: World Robotics 2013 Industrial Robots & Service Robots. http://www.ifr.org/uploads/media/Executive_Summary_WR_2013_01.pdf
4. Lowe D. G. Distinctive Image Features from Scale-Invariant Keypoints // Int. J. of Computer Vision. 2004. V. 60. № 2.
5. Bay H., Tuytelaars T., Van Gool L. SURF: Speeded Up Robust Features // Proc. 9th European Conf. on Computer Vision. Graz, Austria. 2006. V. 3951.
6. Hirschmuller H. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2008. V. 30. № 2.
7. Viola P., Jones M. Robust Real-time Object Detection // Workshop on Statistical and Computation Theories Vision. July, 2001.
8. Dalal N., Triggs B. Histograms of oriented gradients for human detection // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005. V. 1.
9. Ciresan D. C., Meier U., Masci J., Schmidhuber J. Multi-Column Deep Neural Network for Traffic Sign Classification // Neural Networks, 2012.
10. Large Scale Visual Recognition Challenge. www.image-net.org/challenges/LSVRC/2012/.