

**АНАСТАСИЯ ПОПОВА,**

научный сотрудник Института проблем лазерных и информационных технологий Российской академии наук (ИПЛИТ РАН)  
 anastasia.popova@fsmmedia.ru

# ТОЧКА И ПИКСЕЛЬ ЗРЕНИЯ

Машинное (или компьютерное) зрение — программно обусловленный процесс извлечения полезной информации из изображений, и в частности обнаружение, слежение и классификация объектов.

*Посредством глаза, но не глазами  
 смотреть на мир умеет разум.  
 В. Блейк*

Для решения задач детектирования объектов и определения положения в пространстве в течение миллионов лет природа (и, в частности, венец ее творения) использует оптику. Более того, утверждается, что чуть ли не девяносто процентов всей информации вообще человек получает через свою систему детектирования и распознавания образов, затрачивая на обработку данных более четверти всех ресурсов мозга. Логично, что с давних пор человеку хотелось разобраться в том, как он видит. И вполне логично, что, как только человек находит достаточно подтверждений своей теории зрения, он пытается создать такую систему самостоятельно. К тому же считается, что решение задачи машинного зрения на уровне человека равносильно решению задачи искусственного интеллекта. Достойный вызов, иначе не скажешь.

**ВИДИМО-НЕВИДИМО**

Так все же, что такое зрение? Когда мы сможем уверенно сказать, что машина видит? Пожалуй, когда поймем, что из пикселей входных данных извлекаются полезные сведения, не перегруженные лишней информацией. Как и в случае челове-

ческого зрения, этот процесс можно рассматривать как некоторое отображение одного представления в другое, причем исходное представление образуется массивами зарегистрированных значений яркости изображения. С первым этапом, получением входных данных, все более-менее понятно. Несмотря на разную аппаратную реализацию, глаз очень похож на камеру: есть оптическая система и фокусирующая линза-хрусталик, роль диафрагмы играет радужка, а матрицы — фоторецепторы на задней стенке глаза. Ни в коем случае не пытаюсь осудить автора устройства глаза, тем не менее хотелось бы выделить аспекты, которые могли бы быть подвергнуты усовершенствованию. Первое — это спектральная чувствительность глаза. Человеческий глаз способен регистрировать электромагнитное излучение в сравнительно небольшом диапазоне длин волн, названном соответственно видимым излучением. В то время как многие насекомые (и даже некоторые птицы!) обладают возможностью видеть в ультрафиолете, а раки-богомолы — даже в ближнем инфракрасном. Да что там раки-богомолы! Матрица фотоаппарата способна регистрировать излучение инфракрасного спектра, но фильтром *Hot-mirror* эта возможность убирается: иначе снимки будут выглядеть непривычно для челове-

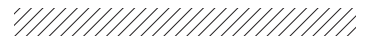
ского глаза... А вот для машинного зрения «видимый» спектр может быть много, много шире. Другой спорный аспект человеческого зрения — неравномерность расположения фоточувствительных элементов. Наибольшая плотность наблюдается на небольшом участке сетчатки, это место носит название «центральная ямка», *fovea*, и именно она предоставляет нам картинку в хорошем разрешении. И почему эта область только одна? У некоторых птиц их по две, или даже по три в каждом глазу, а у собак и кошек это целая зрительная полоса. В результате получается, что в хорошем разрешении мы видим только маленький кусочек пространства, а все остальное размыто, и утверждается, что при таком разрешении мозг не может проводить распознавание. Чтобы это компенсировать, глаз совершает саккады — быстрые согласованные движения, происходящие в одном направлении одновременно. Причем многочисленные эксперименты показали, что порядок осмотра изображения не произволен, а напрямую зависит от цели осмотра. Поэтому то, что мы видим, решающим образом зависит от того, на что мы смотрим: ведь невозможно отслеживать сразу все явления, происходящие в мире, «фовии» не хватит. Мы же не кошки и даже не птицы. А видимая четкость мира — иллюзия, созданная когнитивной системой. Но вот камере никто не мешает воспринимать все и сразу. Лишь бы обработка данных при этом не подкачала.

**УЗРЕТЬ ИСТИНУ**

Переходя к этапу обработки полученной информации, следует отметить в нашей зрительной системе наличие встроенной многоуровневой автономной корректировки искажений. Так устраняются сферическая и хроматическая аберрации, проводится цветокоррекция, устанавливается баланс белого, формируется стереоскопическое изображение. Причем все эти преобразования происходят автоматически, без контроля сознания: практически идеальная адаптивная камера!

Для продолжения разговора о технологии обработки изображений рассмотрим теорию зрения, предложенную гениальным Дэвидом Марром (David Courtnay Marr).





ров, а также задать положительные и отрицательные примеры, а зависимости между этими факторами подбираются уже автоматически, также составляется формула, которая позволяет принимать решения. Бывает, что и сами факторы выделяются автоматически. Результаты работы таких систем довольно интересны, но пока недостаточно точны. Сложно найти применение технологии, допустим, в системах безопасности, при успешном результате лишь в девяноста случаях из ста.

Да и вообще формализация картинок — это довольно сложно. Вспомним слепцов, щупавших слона... Но ведь у «зрячих» машин ситуация лишь ненамного лучше! На трехмерный объект можно смотреть с разных сторон, при разном освещении, в разном окружении... И для машины это будут совершенно разные объекты. Почему же мы можем по плоской картинке восстановить перспективу? Видя сходящиеся к горизонту линии проводов, мы утверждаем, что они параллельны, а глядя на изображение комнаты без единого прямого угла, смело говорим, что между потолком и стеной угол ровно девяносто градусов. Да потому, что в нашем мозгу хранится колоссальный объем информации, и каждую секунду мы распознаем изображения и сравниваем их с виденным ранее. Процесс обучения не останавливается с самого рождения.

А как «объяснить» компьютеру, видеовещу лица исключительно в анфас, что лицо в профиль — это тот же человек? Компьютерное зрение сталкивалось с подобными проблемами очень давно, с момента своего появления, и периодически находило эффективные частные решения. Так, мы можем опознать сдвинутый в сторону предмет, последовательно передвигая свой проверочный паттерн по всему изображению, можем справляться с масштабированными или повернутыми картинками с помощью соответствующих признаков, но эффекты перспективы и поворот предмета в пространстве сцены — похоже, вещи качественно другого уровня. Здесь нам нужно знать, как предмет выглядит со всех сторон, получить его истинную трехмерную форму, иначе нам не с чем работать... Эту проблему можно решать разными способами, например используя видео- или биноккулярную съемку, или даже трехмерный сканер. Но по-прежнему нельзя говорить об общих правилах создания «понимающей» машины, и по-прежнему решения будут свои для каждой конкретной задачи.

### ЧТО Я ВИЖУ?

Нельзя сказать, что машинное зрение — дело далекого будущего. Уже сегодня мы встречаемся с его работой при считывании штрих- и баркодов. Также успешно работают системы

считывания автомобильных номеров, дефектоскопия, распознавание лиц.

Во многих областях науки и техники уже есть четко сформулированные задачи для машинного зрения, решением которых занимаются ведущие исследовательские центры и робототехнические компании.

Большой интерес представляет использование технологий машинного зрения в системах управления промышленными процессами, а именно — в гибком автоматизированном производстве. В условиях быстро меняющихся технологий сложно недооценить использование промышленных роботов: исчезает необходимость переоборудования конвейера, достаточно просто переписать программу робота. Использование технологий машинного зрения позволяет дополнительно расширить диапазон действий робота.

Эти технологии успешно применяются для бортовых компьютеров последних поколений, а также в создании автономных транспортных средств, способных преодолевать сотни километров по дорогам общего пользования, не препятствуя дорожному движению.

Отдельно стоит отметить применение разработок в анализе медицинских изображений. Разумеется, пока что машине далеко до квалифицированного врача, однако использование результата работы программы как дополнительного источника информации может оказаться критически важным. Уж чем механический глаз хорош, так это тем, что он никогда не устает и не «замыливается».

Безусловно, ценными технологии машинного зрения являются для сортировки больших массивов изображений, для организации систем видеонаблюдения. Важную роль играет машинное зрение и в системах дополненной реальности. Ведь чтобы реальность дополнить, нужно понять, что это за реальность и как конкретно ее необходимо дополнять. Также оно может использоваться как изящная система ввода информации в системе человеко-машинного взаимодействия.

Интересное применение машинного зрения — использование смартфона для поиска потерявшегося животного. Компания из Сан-Франциско для этого представляет



бесплатное приложение, основанное на технологии распознавания лиц. В случае пропажи владелец животного загружает его фотографию в базу данных, и там она хранится как шаблон. Любой человек, встретивший на улице бездомного зверя, может сфотографировать его и, воспользовавшись сервисом, узнать, как вернуть его владельцу.

Вообще использование различных технологий компьютерного зрения для мобильных устройств с камерами представляется крайне перспективным. Ведь устройству получается компактным, мобильным, в качестве приемника выступает уже имеющаяся камера, и для начала работы достаточно загрузить соответствующее программное обеспечение.

Как и в области искусственного интеллекта в целом, в машинном зрении вполне успешно решаются локальные, хорошо формализуемые задачи. Но мечта о «видящем» и «понимающем» компьютере живет и не дает покоя тысячам людей. Кто знает, как далека она от реальности?



В 1973 г. был создан робот Фредди-2, он умел собирать из кубиков машинки, у его управляющего компьютера было 384 кбайт памяти. Требовалось запомнить расположение кубиков, а потом, переключившись в режим сборки, манипулировать ими «по памяти». В настоящее время вычислительные мощности увеличились

многократно, а современные фото- и видеосистемы позволяют получать данные в очень хорошем качестве. И поэтому можно смело сказать, что возможности зрячей машины ограничены только нашей фантазией и искусством программирования. ●

*Иллюстрации автора*