Оптический журнал

Биомедицинская оптика

УДК 004.932.2

Применение искусственных нейронных сетей, учитывающих временную динамику, для анализа движения глаз без специального оборудования

© 2018 г. Е. Ю. Малахова*, аспирант; Е. Ю. Шелепин*; Р. О. Малашин, канд. техн. наук*, **

*Институт физиологии им. И.П. Павлова Российской академии наук, Санкт-Петербург
**Университет ИТМО, Санкт-Петербург

E-mail: katerina.malahova@gmail.com

Поступила в редакцию 05.11.2017

Рассмотрена технология определения направления взора на экране монитора посредством анализа снимков, получаемых с видеокамеры, направленной на пользователя, без использования дополнительного оборудования. Для решения этой задачи предложено комбинирование сверточной нейронной сети, извлекающей высокоуровневые признаки изображений, с нейронной сетью долгой краткосрочной памяти, учитывающей временну́ю динамику глазодвигательной активности. Для обучения модели была собрана представительная база данных видеопоследовательностей с эталонной информацией о направлении взгляда. Эксперименты подтвердили, что учет временно́й информации увеличивает точность регистрации направления взгляда.

Ключевые слова: *айтрекинг, искусственные нейронные сети, анализ видеопотока.*

Коды OCIS: *150.0150*

ВВЕДЕНИЕ

Технологии регистрации и анализа движений глаз широко применяются при проведении научных, диагностических и маркетинговых исследований для изучения психофизиологических реакций на предъявляемые зрительные стимулы. Для осуществления исследований необходимо дорогостоящее оборудование, требующее предварительной калибровки для каждого пользователя, а также, зачастую, дополнительных креплений, шлемов и др. Указанные требования приводят к ограничению сферы применения технологии определения направления взора — айтрекинга, например, в системах расширенной реальности и человеко-машинных интерфейсах. В данной работе исследуется возможность реализации указанной технологии с использованием искусственных нейронных сетей и на основе данных, полученных с камеры, встроенной в корпус пользовательского устройства, оснащенного монитором.

МЕТОДЫ АНАЛИЗА ДВИЖЕНИЙ ГЛАЗ

В большинстве современных систем айтрекинга применяются камеры высокого разрешения, фиксирующие движения глаз с использованием подсветки ближнего инфракрасного диапазона. Лучи направляются в зрачок, отражаются от роговицы глаза, вызывая свечение, которое регистрируется камерой. Такой метод носит название «Центр зрачка/отражение от роговицы» (PCCR — *Pupil Centre/Corneal Reflection*). Использование технологии PCCR с обычной камерой невозможно, так как в видимом спектральном диапазоне из-за бликов теряется возможность однозначного разделения зрачка и роговицы глаза. Разработка технологий, не использующих специализированные устройства, существенно расширяет области возможного применения. При этом во многих случаях меньшая точность по сравнению с технологиями PCCR является приемлемой (например, при разработке человеко-машинных интерфейсов).

ПРИМЕНЕНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ АНАЛИЗА ДВИЖЕНИЙ ГЛАЗ

Следует отметить сложность задачи при отсутствии дополнительных устройств помимо камеры. По изображению требуется определить не только положение головы и глаз, но и положение зрачков. Кроме того, при использовании объективов с разными фокусными расстояниями один снимок не позволяет однозначно определить точку пересечения плоскости монитора и направления взора (возможно лишь извлечение угла направления взгляда относительно камеры). В связи со сложностью поставленной задачи перспективным направлением является применение наиболее мощного (в настоящее время) инструмента автоматического анализа изображений — сверточных нейронных сетей (СНС).

Первоначально в задачах айтрекинга применялись полносвязные нейронные сети (многослойные перцептроны), в которые подавались признаки, извлеченные при предварительной обработке данных. В работе [1] авторы впервые предлагают использовать изображение с камеры как входной сигнал в сеть. В других работах подчеркивается важность применения нейронных сетей в реализации доступного решения по интеракции человек–компьютер для людей с физическими ограничениями [2]. Существенным недостатком большинства ранних работ является низкая стабильность метода к изменениям освещения, а также ограничение на допустимые положения пользователя относительно камеры.

С распространением СНС появилась возможность обойти указанные ограничения, поскольку такие архитектуры позволяют формировать внутри себя структурные описания изображения, инвариантные к большому числу различных преобразований.

Процесс распознавания направления взгляда с применением СНС включает несколько основных стадий: получение изображения, обнаружение глаз либо основных лицевых точек — областей интереса, соотнесение положения зрачков со средой (или координатами экрана в частном случае), выполняемое нейронной сетью.

Архитектура СНС в большинстве проектов айтрекинга остается близкой к классическому варианту [3] и представляет собой последовательно чередующиеся сверточные и обобщающие слои [4, 5]. На вход сети подаются извлеченные с помощью сторонних алгоритмов фрагменты изображения — левый и правый глаза, лицо и маска, определяющая положение лица на снимке в отдельных случаях. Такой подход, с точки зрения авторов настоящей статьи, является оправданным, поскольку выделение подзадач позволяет полноценно использовать не только довольно успешные алгоритмы выделения областей лиц и глаз, но и большие базы данных, доступные для их обучения.

В работе [4] предлагается алгоритм, позволяющий в тестовом режиме (после обучения) обрабатывать до 24 кадров в секунду. Для каждого из глаз была обучена отдельная нейронная сеть, определяющая «номер» одной из трех (97% точности) или семи (86% точности) областей экрана. Авторы работы отмечают, что классификация по вертикальной оси более затруднительна, что может быть связано с тем, что при перемещении взгляда вниз зрачок частично прикрыт веком. Несмотря на общую высокую надежность, применение алгоритма [4] в задачах айтрекинга, с нашей точки зрения, весьма ограничено из-за низкой разрешающей способности.

Другим примером является алгоритм iTracker [5], позволяющий обрабатывать 10–15 кадров в секунду и предсказывающий расстояние от камеры до текущего направления взгляда. Проект рассчитан на работу с портативными устройствами и определяет точку фиксации со средней ошибкой в 1,71 см на смартфонах и 2,53 см на экранах планшетов без дополнительной калибровки. Нейронная сеть обучалась на обучающей выборке из 2,5 миллионов кадров, собранных от более чем 1450 человек в процессе пользования портативными устройствами. Авторы старались добиться устойчивой работы с разными пользователями и устройствами (как было сказано выше, это в общем случае является недостижимым из-за недоопределенности задачи). Аналогично другим проектам, из изображений предварительно извлекаются области интереса — в данном случае это левый и правый глаза, лицо и маска, отражающая положение головы относительно снимка целиком. Указанные фрагменты подаются в отдельные подструктуры сети (сверточные подсети), информация из которых затем объединяется на последующих полносвязных слоях, где принимается решение о направлении взгляда.

УЧЕТ ВРЕМЕННЫХ ПАТТЕРНОВ ДВИЖЕНИЯ ГЛАЗ ПРИ РАСПОЗНАВАНИИ НАПРАВЛЕНИЯ ВЗОРА

Все перечисленные выше алгоритмы обрабатывают информацию покадрово, нивелируя важнейшую составляющую глазодвигательной активности, проявляющуюся в динамике данного процесса. Паттерны движения глаз человека являются характерными при решении различных задач, хотя и не тривиальными. Например, при чтении глаза человека двигаются по строчкам текста, хотя и со значительными отклонениями. Предоставление искусственной нейронной сети возможности извлекать и учитывать эту информацию потенциально может помочь существенно повысить

точность прогноза. В качестве возможного решения, учитывающего сведения о предыдущих состояниях, возможно применение 4D свертки, когда в сеть подается не одно изображение, а набор из текущего кадра и определенного количества полученных ранее. Этот подход значительно повышает требования к вычислительным ресурсам, поэтому, по сути, является не применимым к задачам, где требуется учет долгосрочных временных зависимостей. Другим, более экономичным в этом плане, решением является обработка временных последовательностей, состоящих не из «сырых данных», а из высокоуровневых признаков, извлеченных сетью. Такие последовательности могут обрабатываться как сверточными, так и рекуррентными нейронными сетями. Именно такой подход был использован в настоящей работе.

Для проведения экспериментов была собрана база данных, в которую включены записи 32 респондентов, выполнявших шесть видов типичных заданий, характерных при ежедневной работе с компьютером. В нескольких видах заданий пользователи получали широкую свободу действий, тем самым обеспечивалась естественность полученных данных. Суммарный объем видеозаписей составил 19 час. В ходе работы записывались показания айтрекера (*Gazepoint* GP3 с программным обеспечением Gazepoint Analysis), встроенной вебкамеры, а также видеозапись происходящего на экране.

Для учета временных паттернов движения глаз из изображений с помощью обученной СНС iTracker [5] извлекались признаки (активации предпоследнего слоя), которые затем подавались на вход сети с использованием долгой краткосрочной памяти [6]. Расширение рекуррентных архитектур с помощью долгой краткосрочной памяти на настоящий момент является чрезвычайно популярным средством обработки сигналов, где требуется выявление долгих временных зависимостей. Согласно этому подходу нейронная сеть получает возможность фиксировать активность нейрона на неопределенно долгое время посредством управления специальными сигналами для «записи», «чтения», «сохранения» и «стирания». При этом важно, что процесс подачи сигналов так же настраивается в процессе обучения алгоритмом обратного распространения ошибки. Общая архитектура предлагаемого решения, учитывающего временные особенности глазодвигательной активности, представлена на рис. 1. Для сопоставления результатов применения указанного подхода с подходом, где оценка положения взора проводится на основе одного кадра, был также обучен двуслойный перцептрон для адаптации активаций сети iTracker к данным, используемым в работе, так как сеть изначально обучалась на данных с портативных устройств. Модели обучались на 700 тысячах наблюдений (видеокадр и соответствующее ему направление взора на экране), тестирование проводилось на 100 тысячах наблюдений.

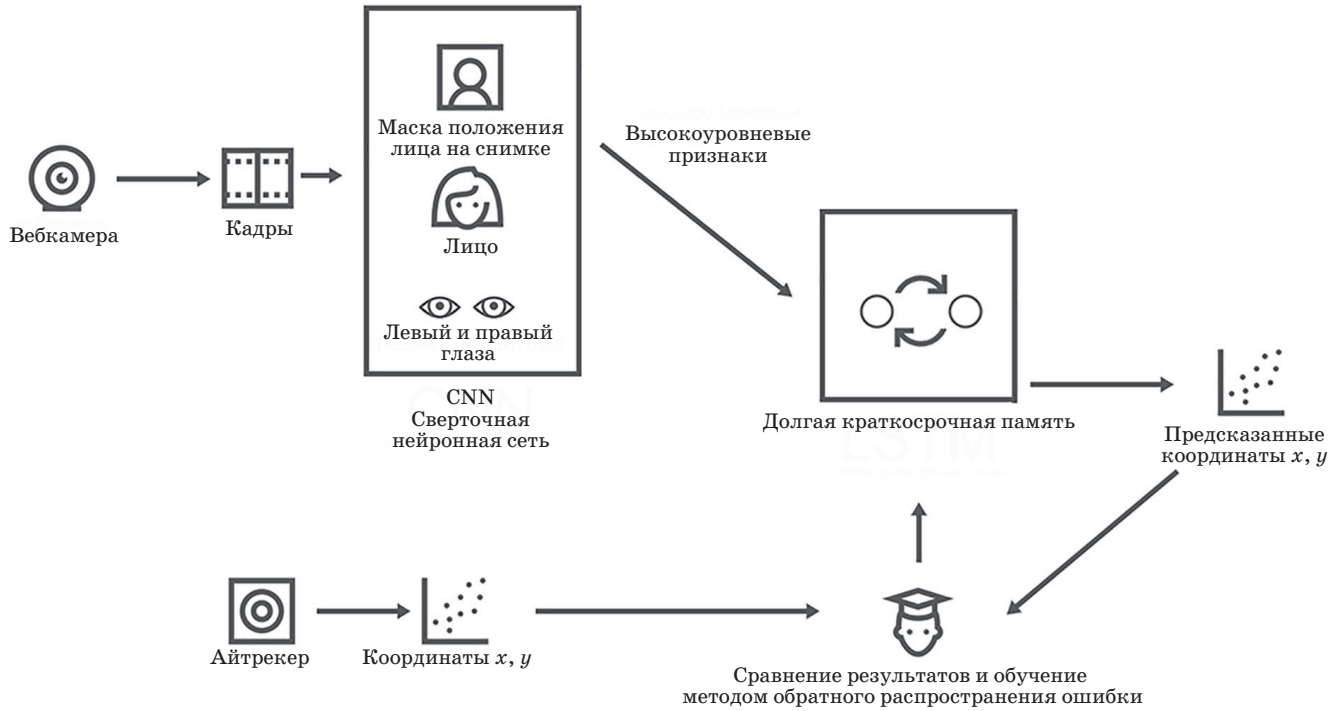

Рис. 1. Общая схема предлагаемого решения.

РЕЗУЛЬТАТЫ

В работе проведено сравнение эффективностей предлагаемой модели (СНС + долгая краткосрочная память) и базовой модели на основе только сверточной сети — получено существенное увеличение точности. Среднее относительное отклонение (с учетом размеров экрана) между предсказанным и эталонным направлением E взора снизилось с 0,29 до 0,26, где E рассчитывается по формуле

$$E = \sqrt{\left(\frac{x - xg}{W}\right)^2 + \left(\frac{y - yg}{H}\right)^2},$$

где W и H — ширина и высота экрана соответственно, (xg, yg) — координаты направления взора, предсказываемые с помощью профессионального оборудования, (x, y) — результат работы нейронной сети. При этом количество предсказаний с относительной ошибкой $E < 0{,}2$ увеличилось на 6% (рис. 2). Корреляция координат по оси Y увеличилась с 0,17 до 0,34.

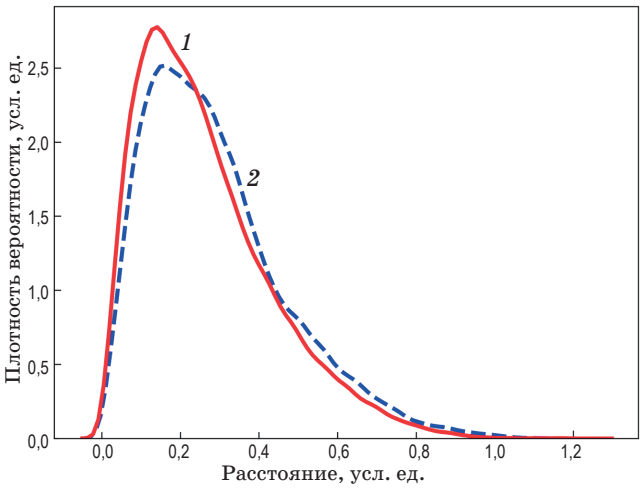

Рис. 2. Результаты учета временно́й информации при предсказании направления взора. Распределение значений по расстоянию между эталонной и предсказанной координатами. *1* — модель СНС + долгая краткосрочная память, *2* — модель СНС + перцептрон.

ЗАКЛЮЧЕНИЕ

В настоящее время решения на основе больших СНС позволяют определять направления взора по изображению в реальном времени. Для учета временно́й динамики глазодвигательной активности было предложено комбинировать СНС со слоями долгой краткосрочной памяти. Согласно экспериментам, учет такой динамики позволяет уменьшить среднюю ошибку восстанавливаемого направления взора на 10%.

ЛИТЕРАТУРА

1. *Sewell W., Komogortsev O.* Real-time eye gaze tracking with an unmodified commodity webcam employing a neural network // ACM. 2010. P. 3739–3744.
2. *Demjén E., Abosi V., and Tomori Z.* Eye tracking using artificial neural networks for human computer interaction // Physiological Research. 2011. V. 60. № 5. P. 841.
3. *Krizhevsky A., Sutskever I., Hinton E.G.* ImageNet classification with deep convolutional neural networks // Advances in Neural Information Proc. Systems. 2012. V. 2. P. 1097–1105.
4. *George A., Routray A.* Real-time eye gaze direction classification using convolutional neural network // SPCOM. 2016. P. 1–5.
5. *Krafka K., Khosla A., Kellnhofer P., Kannan H., Bhandarkar S., Matusik W., Torralba A.* Eye tracking for everyone // Proc. IEEE Conf. on Computer Vision and Pattern Recognition. 2016. P. 2176–2184.
6. *Hochreiter S., Schmidhuber J.* Long short-term memory // Neural Computation. 1997. V. 9. № 8. P. 1735–1780.