

GazeT: улучшение определения трехмерного вектора направления взгляда оператора

М.Д. Красильников^{*,**}, М.Ю. Никитин^{**}

^{*}Московский государственный университет имени М.В. Ломоносова, Москва, Россия

^{**}Tevian, Москва, Россия

Поступила в редколлегию 18.10.2024 г. Принята 09.12.2024 г.

Аннотация—Определение направления взгляда — важная задача в области компьютерного зрения, находящая применение в различных сферах, таких как мониторинг водителей, контроль операторов сложных систем, маркетинг и разработка пользовательских интерфейсов. Существующие методы определения трёхмерного вектора взгляда требуют больших объёмов дорогостоящих данных для обучения нейросетевых моделей. В данной статье предлагается новый подход к улучшению определения трёхмерного вектора направления взгляда оператора, основанный на использовании большого количества данных, собранных для определения двухмерной точки фокусировки взгляда. Созданный набор данных GazeT, включающий более 144000 изображений от 6744 уникальных участников, позволил повысить точность моделей при решении задач определения двухмерной точки фокусировки взгляда, трёхмерного вектора направления взгляда, а также производных задач, таких как классификация направлений взгляда водителей. Результаты экспериментов подтверждают, что предложенный подход позволяет снизить зависимость от дорогостоящих трёхмерных данных и существенно улучшить качество работы алгоритмов определения направления взгляда.

КЛЮЧЕВЫЕ СЛОВА: направление взгляда, краудсорсинг, набор данных, сверточные нейронные сети.

DOI: 10.53921/18195822_2024_24_4_421

1. ВВЕДЕНИЕ

Определение направления взгляда является важной задачей в области компьютерного зрения, которая находит свое применение во множестве сфер. Алгоритмы распознавания направления взгляда активно используются для мониторинга водителей, чтобы снизить риск дорожно-транспортных происшествий, а также в производстве, где необходимо контролировать внимание операторов, управляющих сложными машинами и системами. Технологии определения направления взгляда также востребованы в маркетинге и разработке пользовательских интерфейсов [17]. Они позволяют анализировать, на какие элементы пользователи обращают больше внимания, тем самым улучшая дизайн и эффективность интерфейсов и рекламы.

Существует два основных подхода к решению задачи определения направления взгляда: определение двухмерной координаты точки фокусировки взгляда и предсказание трехмерного вектора направления взгляда в пространстве. В первом случае алгоритм определяет, на какую точку экрана смотрит оператор. Это широко применяется в задачах отслеживания поведения пользователей в интерактивных системах, используется в задачах адаптивного сжатия видео-контента [18]. Определение трехмерного вектора является более общей и в то же время более сложной задачей, поскольку она не ограничена единственным сценарием использова-

ния в отличие от задачи определения точки на экране. Примером может выступать задача классификации взгляда водителей [4].

Однако, задача предсказания трехмерного вектора взгляда сталкивается с рядом проблем. Ключевая из них — необходимость больших объемов данных для обучения моделей на основе нейронных сетей. Сбор таких данных в трехмерном пространстве требует использования сложных и дорогостоящих устройств для точной регистрации положения головы и глаз человека, что делает процесс длительным и финансово затратным. На рисунке 1 можно увидеть примеры конструкций для сбора таких данных. С другой стороны, сбор данных для определения двухмерной точки на экране значительно проще и может быть выполнен с использованием обычных веб-камер и краудсорсинговых платформ, этот подход использовался для сбора набора данных GazeCapture [7]. Это открывает возможность создания масштабных и разнообразных наборов данных с небольшими затратами, что является критически важным для дальнейшего обучения и совершенствования нейросетевых моделей.

Цель данной работы — улучшить качество работы алгоритма определения трехмерного вектора направления взгляда, используя данные, собранные для определения двухмерной точки фокусировки взгляда. Предлагается метод, в котором большое количество данных с разметкой для определения двухмерных точек, полученных через краудсорсинг, используется для предобучения нейросетей, что позволяет улучшить точность предсказаний в более сложных задачах трехмерного анализа. Этот подход помогает снизить зависимость от дорогостоящих трехмерных данных, что делает разработку систем определения взгляда более доступной и масштабируемой.

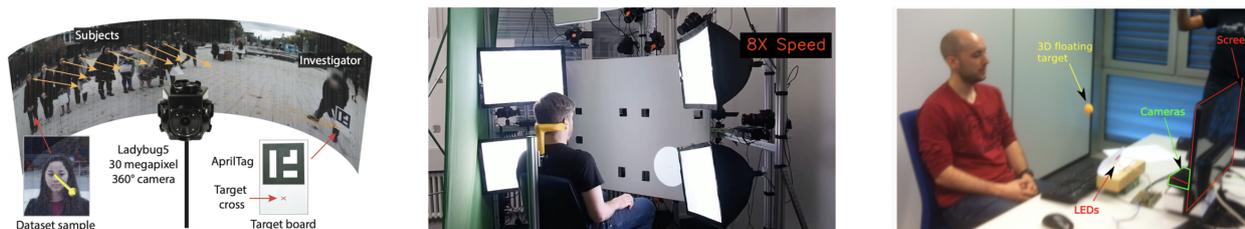


Рис. 1. Сбор данных для задачи определения трехмерного вектора. EYEDIAP [13], Gaze360 [12], ETH-XGaze [11]

Новизна данной работы заключается в следующем:

1. Разработан метод сбора больших объемов данных для задачи определения двухмерной координаты точки фокусировки взгляда с использованием краудсорсинговых платформ. Были решены проблемы, которые могут возникнуть при сборе через краудсорсинговую платформу.
2. С помощью разработанного метода был получен набор данных и был выложен в открытый доступ¹
3. Предложен подход к использованию двухмерных данных для улучшения качества предсказаний в задаче определения трехмерного вектора направления взгляда.
4. Показано, что добавление таких данных в обучение позволяет улучшить качество работы методов определения трехмерного вектора направления взгляда и других производных задач на примере классификации зон взгляда водителей.

¹ https://github.com/mirmashe1/gazet_dataset

2. ОБЗОР ЛИТЕРАТУРЫ

2.1. Методы определения направления взгляда

Классические методы отслеживания направления взгляда часто основываются на геометрических моделях глаза и традиционных алгоритмах машинного обучения. Например, в статье [1] предложена геометрическая модель глаза, которая описывает его как треугольник, и применяется алгоритм Viola Jones для обнаружения лица и глаз. Другие работы, такие как [8], также используют некоторые предположения о форме глаза, и полагаются на методы оптимизации для точного определения направления взгляда. Однако традиционные методы имеют ограничения по сравнению с современными нейросетевыми подходами, особенно в сложных условиях съемки, например, при наличии изменяющегося освещения и перекрытий лица.

Для более точного решения задачи определения двухмерной точки фокусировки взгляда используются глубокие нейронные сети. В работе [2] предложен метод с использованием CNN для классификации направления взгляда, что привело к значительному улучшению точности по сравнению с традиционными методами. Метод ITracker [7] использует нейронную сеть, которая принимает на вход четыре изображения, включая маску расположения лица в кадре и изображения глаз по отдельности. Предложенный в статье [14] метод использует технику обучения с подкреплением для персонализации модели, что позволяет адаптировать ее для конкретного пользователя непосредственно перед использованием.

Для задач определения трехмерного вектора взгляда применяются более сложные модели. Например, в работе [5] предложен метод с использованием трех входных изображений — обоих глаз и лица, что позволяет адаптировать веса для каждого из них и улучшить точность предсказания направления взгляда при зашумленности изображений. В статье [9] предложены разреженные сверточные операции для извлечения признаков из изображения глаза, что улучшает оценку направления взгляда в условиях сложного освещения. В работе [6] авторы предложили модель, которая получает и совмещает вероятностные оценки направления взгляда отдельно с изображений глаза и сцены. Такой подход делает модель устойчивой к перекрытиям. В статье [10] предложена адаптивная нейросеть с двумя модулями — для грубой оценки и точной настройки направления взгляда, что повысило производительность и адаптивность модели.

В некоторых задачах определение направления взгляда используется как компонент более сложных систем. Например, в статье [4] авторы использовали алгоритмы регрессии и классификации направления взгляда для предсказания поведения водителей. В статье [17] методы определения направления взгляда использовались для улучшения восприятия информации в образовательных и маркетинговых целях, а статья [18] описывает использование полуавтоматического моделирования визуального внимания для адаптивного сжатия видео, что позволило улучшить эффективность передачи данных.

2.2. Существующие наборы данных для задачи определения направления взгляда

Наборы данных для задачи определения двухмерной точки взгляда на экране обычно проще собирать по сравнению с трехмерным взглядом. В статье [3] предложен способ автоматической разметки данных для обучения моделей классификации зоны, на которую смотрит оператор. Набор данных GazeCapture, собранный для метода ITracker [7], использовал краудсорсинговые платформы для получения данных. Исполнителям необходимо было нажимать на точку на экране, и в этот момент делалась их фотография вместе с координатами нажатия. Такой подход позволяет получать большое количество данных с небольшими затратами. Еще один набор данных — MPIIGaze [15], был собран с участием 15 человек, которые работали в различ-

ных условиях в течение 40 дней, что позволило зафиксировать различные сценарии и условия окружающей среды.

Сбор размеченных данных для задачи трехмерного определения направления взгляда гораздо сложнее и требует точных параметров сцены или специального оборудования. Примеры систем для сбора таких данных можно увидеть на рисунке 1. Например, для создания набора данных EYEDIAP [13] использовались синхронизированные камеры RGB и RGB-D с LED лампами для точного захвата всех аспектов сцены. Другие наборы, такие как ETH-XGaze [11], включают данные с экстремальными углами поворота головы, что позволяет моделям анализировать взгляд в сложных условиях. В статье [12] представлен набор данных Gaze360, собранный в естественных условиях с использованием панорамных камер, что позволило одновременно записывать несколько людей в одном кадре и учитывать разнообразные сценарии.

3. МЕТОД СБОРА ДАННЫХ ДЛЯ ЗАДАЧИ ОПРЕДЕЛЕНИЯ ДВУХМЕРНОЙ КООРДИНАТЫ ТОЧКИ ФОКУСИРОВКИ ВЗГЛЯДА

Сбор данных для задачи определения трёхмерного вектора направления взгляда является сложной и дорогостоящей задачей. Для этого требуется точная настройка сцены или специализированное оборудование, а также участие заинтересованных людей, что делает такие выборки трудными для масштабирования. Примеры конструкций для сбора таких данных можно увидеть на рисунке 1. С другой стороны, получение данных для задачи определения двухмерной точки является менее требовательной задачей, для этого достаточно иметь персональный компьютер и веб-камеру, так как для получения такой разметки нужно всего лишь фиксировать координаты точки нажатия [7]. Такой подход позволяет использовать краудсорсинговые платформы для сбора данных и легко масштабировать этот процесс.

Особенностями сбора данных с помощью краудсорсинговых платформ являются небольшая стоимость итогового набора данных, большое количество уникальных исполнителей, высокая скорость получения данных, но в то же время низкая заинтересованность исполнителей в итоговом качестве разметки. Основная задача заключается в том, чтобы сформировать задания так, чтобы их выполнение можно было контролировать автоматически или с минимальными расходами.

В данной работе предлагается метод сбора данных для задачи определения двухмерной координаты фокусировки взгляда оператора с помощью краудсорсинговой платформы. Этот подход позволяет эффективно контролировать процесс разметки, минимизируя ошибки и сокращая затраты. Далее описан сам процесс и методы контроля качества собранных данных.

3.1. Сервис для сбора данных

Для сбора данных был разработан веб-сервис, ссылка на который выдавалась исполнителям через краудсорсинговую платформу Яндекс.Толока. При заходе на страницу задания исполнители должны были следовать инструкциям, включающим проверку работы веб-камеры, размещение камеры по центру сверху монитора и указание размера диагонали экрана. Интерфейс этапа настройки можно увидеть на рисунке 2. Эти параметры в дальнейшем играли важную роль при вычислении координат. После ввода параметров исполнители приступали к основному заданию. Задание заключалось в том, чтобы нажимать на серию последовательно появляющихся точек. В момент нажатия на каждую точку веб-камера делала снимок и пара «фотография – координата нажатия» сохранялась на сервере. Схематично это показано на рисунке 3. Авторы работы [19] показывают, что измерение визуального внимания через нажатия на экране имеет высокую корреляцию с традиционными методами измерения с помощью приборов для отслеживания глаз. Выводы из этой работы позволяют принять, что при нажатии на точку на экране человек будет смотреть именно на нее.

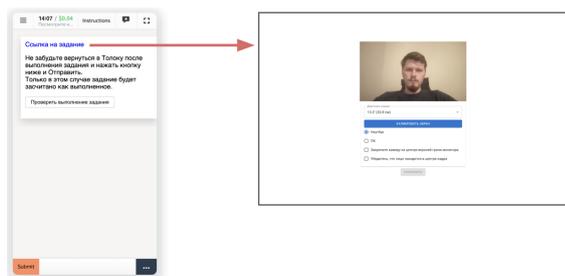


Рис. 2. Этап настройки.

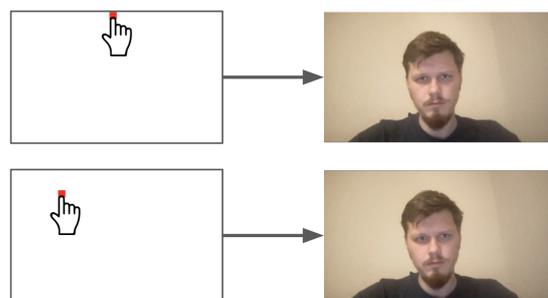


Рис. 3. Демонстрация процесса получения данных.

3.2. Проблемы и решения

В процессе сбора данных через краудсорсинговую платформу возникли определенные трудности, которые требовали внимательного подхода к проверке качества собранных данных. Часто встречалось отсутствие лица на изображениях или использование статичных фотографий вместо видеопотока с камеры. Для решения этой задачи была реализована автоматическая проверка изображений с помощью детектора лиц. Если детектор не находил лицо или все изображения сессии оказывались идентичными, то такие данные исключались из анализа.

Кроме того, некоторые участники намеренно вводили неправильные параметры, такие как размер диагонали экрана, что приводило к ошибкам в расчётах и невозможности расчета координат в сантиметрах. Для борьбы с этим была разработана схема проверки корректности указанной диагонали, где исполнителю предлагалось с одной стороны указать диагональ ручную, а с другой стороны отметить 10 сантиметров на экране с помощью линейки и ползунка на экране. Размер экрана в пикселях получался автоматически. Используя эту информацию можно получить рассчитанную диагональ. Эта диагональ сравнивается с той, что была выбрана вручную. Если они не сходятся, то исполнитель где-то ошибся, и ему предлагалось это исправить. Такой подход обеспечил проверку корректности ввода диагонали еще до начала разметки и без ручной проверки. На рисунке 4 можно увидеть влияние этой проверки на распределение диагоналей в полученных сессиях. Можно заметить, что полностью исчезли экстремально малые и большие размеры экранов, что говорит об эффективности введенной проверки.

Также было замечено, что участники не всегда правильно размещали камеру, что нарушало консистентность разметки собираемых данных. Для выявления этого отклонения был добавлен контрольный элемент: одна из точек для нажатия располагалась в верхней части экрана, и при её нажатии исполнитель будет смотреть практически в камеру. Таким образом, при пост-проверке можно было быстро разметить все сессии на предмет правильного расположения камеры, отвечая на вопрос «смотрит ли человек на этом фото вам в глаза?» и исключить сессии с нарушениями. Таким образом было исключено 723 сессии, то есть около 10% выборки. На рисунке 5 показано визуальное различие между правильным и неправильным расположением камеры.

3.3. Итоговый набор данных GazeT

В результате был собран набор данных GazeT, который включает 6744 сессий из более 144000 фотографий. Благодаря использованию краудсорсинга, GazeT содержит больше уникальных участников и более разнообразные сцены, чем многие другие наборы данных. А с помощью предложенных методов валидации сессий получилось исключить много некоррект-

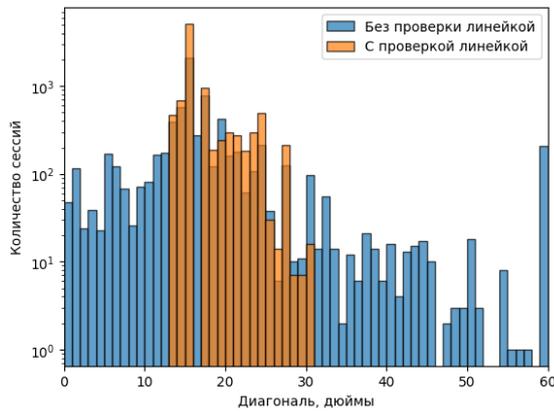


Рис. 4. Сравнение распределения диагоналей при использовании предложенного способа проверки диагоналей.



Рис. 5. Проблема расположения камеры не по центру.

ных сессий и некорректных диагоналей экрана. Сравнение характеристик наборов данных можно увидеть в таблице 1.

Название	Количество уникальных человек	Количество фото	Трехмерный вектор направления взгляда	Двухмерная точка фокусировки взгляда
MPIIGaze [15]	15	~213 тыс.	+	+
Gaze360 [12]	238	~100 тыс.	+	-
ETH-XGaze [11]	110	~1 млн.	+	+
EYEDIAP [13]	16	~85 тыс.	+	+
GazeCapture [7]	~1,5 тыс.	~2,5 млн.	-	+
GazeT	~6,7 тыс.	~144 тыс.	-	+

Таблица 1. Сравнение открытых наборов данных и полученного в рамках этой работы GazeT.

4. ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА

В проведенных экспериментах применялся подход из метода ITracker [7]. Первоначально с помощью детектора лица происходит локализация лица и области глаз. Направление взгляда регрессируется нейронной сетью, принимающей изображение области глаз и маску расположения лица в кадре. Изображение глаз обрабатывается сверточной нейронной сетью, предназначенной для извлечения низко- и высокоуровневых признаков. Маска лица передается в полносвязную нейронную сеть. Конечные выходы ветвей объединяются и подаются в полносвязную сеть для регрессии направления взгляда. В качестве основной сети для извлечения признаков использовалась архитектура ResNet10, основанная на архитектуре residual network [16], которая предварительно обучалась на задаче распознавания лиц, а полученные веса были использованы для инициализации в задаче определения направления взгляда.

4.1. Экспериментальная оценка для задачи определения точки фокусировки взгляда оператора на экране

Таблица 2 демонстрирует преимущество использования набора данных GazeT для обучения моделей, занимающихся определением двухмерных точек фокусировки взгляда. Модель

ResNet10, обученная на GazeT, показала значительное улучшение точности как на самом наборе данных GazeT, так и на MPIIGaze, достигнув ошибки 4,04 см и 3,81 см соответственно. Это свидетельствует о том, что обучение на GazeT, содержащем значительно большее количество уникальных людей не только обеспечивает высокую точность на самом наборе данных, но и способствует лучшей обобщающей способности модели на других данных. Этот эксперимент также показывает важность наличия разнообразия лиц в выборках для задачи определения взгляда.

Название	Набор данных для обучения	GazeT Ошибка, см	MPIIGaze Ошибка, см
CA-Net [10]	MPIIGaze	-	4,90
ITracker [7]	MPIIGaze	-	7,67
Gaze360 [12]	MPIIGaze	-	4,66
ResNet10	MPIIGaze	7,07	4,54
ResNet10	GazeT	4,04	3,81

Таблица 2. Сравнение качества работы моделей определения двухмерных точек фокусировки взгляда на экране.

4.2. Экспериментальная оценка для задачи определения направления взгляда оператора

В таблице 3 можно увидеть преимущества использования набора данных GazeT для предобучения моделей определения трехмерных векторов направления взгляда. Наиболее заметное улучшение наблюдается при сравнении модели ResNet10, предобученной на GazeT и обученной на MPIIGaze, с остальными моделями. Эта модель достигла наименьшей ошибки на тестовом наборе MPIIGaze (3,16°), что указывает на явное преимущество предварительного обучения на GazeT в плане повышения точности. Также снизилась ошибка при тестировании такой модели на наборе данных Gaze360 с 23,06° до 21,47°.

Также видно, что предварительное обучение на GazeT значительно улучшает способность модели обобщать информацию и повышает качество предсказания в сравнении с моделью ResNet10, которая не использовала предобучение. В частности, при обучении на наборе Gaze360, использование предобучения на GazeT позволило модели снизить ошибку с 9,90° до 9,54° на наборе Gaze360 и с 8,66° до 7,95° на наборе данных MPIIGaze, что подтверждает ценность предварительного обучения на GazeT. Здесь не видно такого же улучшения, как в случае дообучения на MPIIGaze, и это объясняется сильно различающимся доменом данных.

Таким образом, обучение и предобучение на GazeT обеспечивают существенное повышение точности и универсальности моделей, позволяя добиться более низких ошибок как на MPIIGaze, так и на Gaze360. Это подтверждает, что использование GazeT в качестве предварительного этапа обучения эффективно повышает производительность моделей в задаче определения направления взгляда.

Название	Набор данных для предобучения	Набор данных для обучения	MPIIGaze Ошибка, °	Gaze360 Ошибка, °
CA-Net [10]	-	MPIIGaze или Gaze360	4,42	11,20
ITracker [7]	-	MPIIGaze или Gaze360	7,33	-
Gaze360 [12]	-	MPIIGaze или Gaze360	4,06	11,04
ResNet10	-	MPIIGaze	3,81	23,06
ResNet10	GazeT	MPIIGaze	3,16	21,47
ResNet10	-	Gaze360	8,66	9,90
ResNet10	GazeT	Gaze360	7,95	9,54

Таблица 3. Сравнение качества работы моделей определения трехмерных векторов направления взгляда.

4.3. Экспериментальная оценка для задачи классификации направления взгляда оператора

Таблица 4 показывает существенное влияние предобучения на наборе данных GazeT на точность классификации направления взгляда водителей на наборе данных DriveDS [4]. Модель ResNet10, которая была предобучена на GazeT и затем обучена на MPIIGaze, достигла точности 90,33%, что значительно превышает точность модели, обученной только на MPIIGaze без предобучения (85,32%). Это подтверждает, что предварительное обучение на GazeT позволяет модели лучше извлекать важные особенности, связанные с направлением взгляда, и в дальнейшем более эффективно адаптироваться к задаче классификации направления взгляда на наборе данных MPIIGaze.

Название	Набор данных для предобучения	Набор данных для обучения	Точность
ResNet10	-	MPIIGaze	85,32%
ResNet10	GazeT	MPIIGaze	90,33%

Таблица 4. Результаты работы алгоритма классификации направления взгляда водителей с разными моделями определения направления взгляда.

5. ЗАКЛЮЧЕНИЕ

В представленной работе была описана методика сбора данных для задачи определения двухмерной точки фокусировки взгляда через краудсорсинговую платформу. Были выявлены и решены проблемы, связанные с особенностями работы с исполнителями на краудсорсинговых платформах. С помощью этого подхода был собран большой и разнообразный набор данных GazeT, содержащий в себе 144 тысяч фотографий 6,7 тысяч уникальных людей. Было показано, что добавление этих данных в обучение моделей улучшает точность предсказаний для задачи двухмерной точки фокусировки, задачи определения трехмерного вектора направления взгляда, а также производных задач.

СПИСОК ЛИТЕРАТУРЫ

1. Jabber N. H., Hashim I. A. Robust Eye Features Extraction Based on Eye Angles for Efficient Gaze Classification System //2018 Third Scientific Conference of Electrical Engineering (SCEE). – IEEE, 2018. – С. 13-18.
2. George A., Routray A. Real-time eye gaze direction classification using convolutional neural network //2016 International Conference on Signal Processing and Communications (SPCOM). – IEEE, 2016. – С. 1-5.
3. Wu X. et al. Appearance-based gaze block estimation via CNN classification //2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP). – IEEE, 2017. – С. 1-5.
4. Попов А., Shakhuro V., Konushin A. Operator's Gaze Direction Recognition //Графикон-конференции по компьютерной графике и зрению. – 2021. – Т. 31. – С. 475-487.
5. Zhou X. et al. Learning a 3D Gaze Estimator With Adaptive Weighted Strategy //IEEE Access. – 2020. – Т. 8. – С. 82142-82152.
6. Park S., Spurr A., Hilliges O. Deep pictorial gaze estimation //Proceedings of the European Conference on Computer Vision (ECCV). – 2018. – С. 721-738.
7. Krafska K. et al. Eye tracking for everyone //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 2176-2184.
8. Vicente F. et al. Driver gaze tracking and eyes off the road detection system //IEEE Transactions on Intelligent Transportation Systems. – 2015. – Т. 16. – No. 4. – С. 2014- 2027.

9. Chen Z., Shi B. E. Appearance-based gaze estimation using dilated-convolutions //Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part VI. – Cham : Springer International Publishing, 2019. – C. 309-324.
10. Cheng Y. et al. A coarse-to-fine adaptive network for appearance-based gaze estimation //Proceedings of the AAAI Conference on Artificial Intelligence. – 2020. – T. 34. – №. 07. – C. 10623-10630.
11. Zhang X. et al. Eth-xgaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation //Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16. – Springer International Publishing, 2020. – C. 365-381.
12. Kellnhofer P. et al. Gaze360: Physically unconstrained gaze estimation in the wild //Proceedings of the IEEE/CVF International Conference on Computer Vision. – 2019. – C. 6912-6921.
13. Funes Mora K. A., Monay F., Odobez J. M. Eyediap: A database for the development and evaluation of gaze estimation algorithms from rgb and rgb-d cameras //Proceedings of the symposium on eye tracking research and applications. – 2014. – C. 255-258.
14. He J. et al. On-device few-shot personalization for real-time gaze estimation //Proceedings of the IEEE/CVF international conference on computer vision workshops. – 2019. – C. 0-0.
15. Zhang X. et al. Mpiigaze: Real-world dataset and deep appearance-based gaze estimation //IEEE transactions on pattern analysis and machine intelligence. – 2017. – T. 41. – №. 1. – C. 162-175.
16. He K. et al. Deep residual learning for image recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – C. 770-778.
17. Rodríguez A., Bárcenas E., Molero-Castillo G. Model checking for gaze pattern recognition //2019 International Conference on Electronics, Communications and Computers (CONIELECOMP). – IEEE, 2019. – C. 170-175.
18. Gitman Y. et al. Semiautomatic visual-attention modeling and its application to video compression //2014 IEEE international conference on image processing (ICIP). – IEEE, 2014. – C. 1105-1109.
19. Egner S. et al. Attention and information acquisition: Comparison of mouse-click with eye-movement attention tracking //Journal of Eye Movement Research. – 2018. – T. 11. – №. 6.

GazeT: improved estimation of the three-dimensional vector of the operator's gaze direction

M.D. Krasilnikov, M.U. Nikitin

Determining gaze direction is an important task in computer vision, with applications in areas such as driver monitoring, operator oversight of complex systems, marketing research, and user interface development. Current methods for estimating the three-dimensional gaze vector require large amounts of costly data to train neural networks. In this article, we introduce a new method that improves the estimation of three-dimensional gaze orientation by utilizing extensive data collected for analyzing the two-dimensional gaze focus point. Our new dataset, GazeT, comprising over 144,000 images from 6,744 unique participants, enhances model accuracy in tasks related to determining the two-dimensional gaze focus, estimating the three-dimensional gaze vector, and other related tasks like driver gaze direction classification. Experimental results demonstrate that our approach reduces dependency on expensive three-dimensional data and significantly improves the effectiveness of gaze direction estimation algorithms.

KEYWORDS: gaze direction, crowdsourcing, dataset, convolutional neural networks.