

ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ КОМПЬЮТЕРНОГО ЗРЕНИЯ В УПРАВЛЕНИИ КОМПЬЮТЕРОМ

Яцковец М.А., Завалей В.А.

*Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Республика Беларусь*

Научный руководитель: Скиба И.Г. – магистр, ассистент кафедры ЭВМ

Аннотация: В данной статье анализируются перспективы использования компьютерного зрения для управления ПК. Рассматриваются ключевые методы взаимодействия (жестовое управление, трекинг головы и глаз), реализуемые на базе стандартных веб-камер. Выявляются основные технические и эргономические барьеры внедрения.

Ключевые слова: нейросети, компьютерное зрение, человеко-машинное взаимодействие

Введение. Массовое распространение технологий компьютерного зрения в мобильных приложениях наглядно демонстрирует её зрелость. Даже на устройствах с ограниченными ресурсами успешно реализован надежный трекинг головы и жестов в реальном времени, что доказывает преодоление основных технических барьеров для внедрения подобных решений. Однако наблюдается существенный разрыв между использованием технологий в развлекательных целях и их внедрением в более профессиональную среду персональных компьютеров. В данной работе исследуются перспективы внедрения компьютерного зрения в управление персональным компьютером.

Основная часть. Современное компьютерное зрение предлагает три ключевых способа взаимодействия и регистрации действий. Каждый из них базируется на отслеживании специфических точек тела человека и интерпретации их движений.

Наиболее естественным способом ввода является перенос моторики рук в цифровое пространство. Современные алгоритмы (например, архитектуры, используемые в решениях MediaPipe или OpenPose) позволяют строить 3D-модель кисти по RGB-изображению, отслеживая ключевые точки, такие как суставы пальцев. Примером способа ввода может быть позиция указательного пальца, интерпретируемая как курсор, а «щипок» как триггер клика. В качестве технологической базы используются легковесные сверточные нейросети, обученные на синтетических данных, специально подобранных для повышения устойчивости к перекрытиям, когда пальцы закрыты преградой. В массовом ПО распознавание рук ограничивается определением простейших жестов (например, «палец вверх») для активации эмодзи-реакций в платформах видеосвязи.

Навигация на основе положения головы позволяет использовать голову пользователя как указатель, так и аналог джойстика. Алгоритмы решают задачу Perspective-n-Point, определяя ориентацию головы в пространстве на основе 2D-проекции лица. Как дополнительные средства управления, вертикальные наклоны головы можно использовать для скроллинга текста, горизонтальные – для переключения вкладок или окон. Для команд нажатия клавиш можно использовать жесты лица: поднятие бровей, моргание или открытие рта. Это реализуется через анализ геометрии лицевой сетки, содержащей множество опорных точек лица, или через векторную оценку положения головы в пространстве [1]. Для обычного потребителя использование наклонов и поворотов головы сосредоточено преимущественно в развлекательном сегменте приложений для управления масками или игровыми механиками, практически не применяясь в более серьезных задачах.

Отслеживание положения глаз позволяет определять точку фиксации взгляда на экране. В данном случае взгляд может использоваться для быстрого перемещения курсора. Метод основывается на том, что нейросети сегментируют область радужной оболочки и зрачка, а затем сопоставляют их положение с вектором головы для вычисления финальной координаты

на плоскости монитора [2]. Технологии отслеживания взгляда на данный момент применяются в основном для отслеживания внимания пользователя.

Изначально для точного трекинга тела требовались дорогостоящие системы с нателными датчиками и маркерами, доступные лишь специализированным компаниям. В то же время для отслеживания взгляда требуются высокоточные и дорогостоящие инфракрасные датчики. Большим скачком вперед стало появление сенсоров глубины (например, Microsoft Kinect), которые позволили отказаться от надеваемой электроники, но все еще требовали покупки отдельного устройства. Сегодня, благодаря прогрессу в нейросетях, мы переходим к использованию обычных камер, встроенных в каждый современный ноутбук. Современные алгоритмы способны вычислять глубину и геометрию движений, позволяя реализовать качественное управление на любой веб-камере.

Несмотря на доступность алгоритмов, массовое внедрение систем компьютерного зрения в качестве основного инструмента ввода ограничивается рядом эргономических и технических факторов. В отличие от нажатия физической кнопки, взгляд или положение головы всегда активны. Системе часто бывает сложно отделить случайное действие от осознанной команды. Это требует разработки сложных алгоритмов фильтрации и контекстного понимания, чтобы избежать непреднамеренных взаимодействий. Длительное использование средств управления, связанных с активным движением, быстро вызывает мышечную усталость. Поэтому решения, основанные на жестах и движениях головой, непригодны для продолжительной работы. Они больше подходят для коротких сессий или специфических задач. Обычные веб-камеры чувствительны к освещению, поэтому неправильно поставленный свет и низкая освещенность снижают точность определения ключевых точек, что ведет к недопустимым проблемам с управлением.

Заключение. Методы компьютерного зрения имеют огромный потенциал в задачах облегчения и расширения взаимодействия человека и компьютера. Стремительное совершенствование нейросетевых архитектур и аппаратного обеспечения, вероятно, позволит в ближайшей перспективе минимизировать технические сложности, обеспечив стабильную работу алгоритмов в различных условиях эксплуатации. Это создаст возможности для массовой интеграции бесконтактного управления в программное обеспечение широкого профиля в качестве эффективного дополнения к традиционным средствам ввода.

Социально-экономическая значимость в данном контексте также крайне велика. Использование стандартных веб-камер для отслеживания взгляда и движений обеспечит полноценный доступ к компьютерам для разнообразного спектра пользователей.

Список литературы

1. Fine-Grained Head Pose Estimation Without Keypoints / N. Ruiz, E. Chong and J. M. Rehg // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018
2. MPIIGaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation / X. Zhang, Y. Sugano, M. Fritz, A. Bulling [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/1711.09017> – Дата доступа: 13.02.2026

UDC 004.5

PROSPECTS FOR USING COMPUTER VISION IN COMPUTER CONTROL

Yatskovets M.A., Zavalei V.A.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Skiba I.G. – Assistant of the department of ECS

Annotation: This article analyzes the potential for using computer vision for PC control. Key interaction methods (gesture control, head and eye tracking) implemented using standard webcams are discussed. The main technical and ergonomic barriers to implementation are identified.

Keywords: neural networks, computer vision, human-machine interaction