Ученые СПб ФИЦ РАН научили нейросеть распознавать управляющие жесты на видео для бесконтактного управления устройствами

23.03.2026



Исследователи Санкт-Петербургского федерального исследовательского центра РАН (СПб ФИЦ РАН) разработали интеллектуальный программный комплекс, который распознает управляющие жесты по видео. Такое решение можно использовать для дистанционного управления цифровыми системами и «умной» электроникой. Используемый набор алгоритмов искусственного интеллекта обеспечивает более высокую точность по сравнению с существующими коммерческими аналогами.

Сегодня технологии искусственного интеллекта (ИИ) активно внедряются в самые разные сферы жизни человека. ИИ-сервисы ускоряют обработку больших объемов данных, повышают точность решений, оптимизируют процессы и открывают новые возможности для бизнеса, науки и повседневной жизни.

Одна из важных задач на пути дальнейшего внедрения ИИ-сервисов в жизнь человека — научить системы корректно распознавать весь спектр разнородной информации от пользователя: не только текст и речь, но также мимику, жесты и другие невербальные проявления.

В частности, одним из важных направлений, над которой в России работают специалисты крупнейших IT-компаний, является распознавание управляющих жестов (кивок, большой палец, отрицательное покачивание головой и прочие), которые активно используются в

повседневной жизни, при переговорах и организации работ. Это особенно важно в случае интеллектуальных систем, где, например, случайное нажатие реальной кнопки может повлиять на безопасность работы или в условиях работы человека в виртуальной реальности. Поэтому сейчас подобные решения активно внедряются в коммерческое программное обеспечение с использованием технологий ИИ.

«Мы разработали программное обеспечение, которое позволяет по видео распознавать широкий спектр управляющих жестов, которые в повседневной жизни и на работе активно используют каждый человек. Условно говоря, когда пользователь захочет позвонить или поставить лайк под фото в соцсети, ему достаточно показать в камеру большой палец. Кроме того, в медицине или на пищевом производстве дистанционное управление оборудованием при помощи жестов поможет обеспечить высокий уровень гигиеничности», — поясняет старший научный сотрудник Лаборатории речевых и многомодальных интерфейсов СПб ФИЦ РАН Дмитрий Рюмин.

Разработанное учеными СПб ФИЦ РАН ПО автоматически распознает 34 наиболее часто используемых управляемых жеста (включая отсутствие жеста), среди которых можно, например поставить “лайк”, позвонить по телефону, поставить точку, выбрать предмет, а также знак «rock» (принятый в музыкальной рок-культуре). Достаточно включить ПО на ноутбуке или компьютере и показать жест на камеру.

Масштабный корпус для обучения распознавания жестов содержал более 1 млн изображений людей различного пола, возраста и национальности. Эти данные находятся в открытом доступе для исследователей всех стран мира.

Программное обеспечение исследователей действует в несколько этапов. Система сначала определяет на картинке человека, затем находит его руки. Для повышения качества распознавания используется сравнительно новая технология — нейросетевая модель, которая позволяет получать трехмерную карту глубины изображения. Она помогает определять жесты даже в условиях, когда фон сливается с руками человека. Эта технология позволила добиться рекордной точности распознавания жестов — более 99,6% — по сравнению с существующими аналогами.

«Наша система может использоваться как самостоятельный интерфейс управления цифровыми помощниками, а также как компонент мультимодальных приложений, которые взаимодействуют с пользователем через речь, текст, эмоциональные проявления и другие каналы», — отмечает стажер-исследователь СПб ФИЦ РАН Сергей Федчин.

Исследование поддержано грантом РНФ (№ 24-71-00083). Программное обеспечение получило свидетельство о государственной регистрации.