
Социально-экологические проблемы устойчивого развития

SOCIAL AND ENVIRONMENTAL PROBLEMS OF SUSTAINABLE DEVELOPMENT

УДК 004.932.2+004.582

СИСТЕМА КОМПЬЮТЕРНОГО ЗРЕНИЯ В МОБИЛЬНЫХ УСТРОЙСТВАХ

С. В. ТКАЧЕНКО¹⁾, Т. В. СМИРНОВА¹⁾

¹⁾Международный государственный экологический институт им. А. Д. Сахарова,
Белорусский государственный университет,
ул. Долгобродская, 23/1, 220070, г. Минск, Беларусь

Представлены результаты проектирования системы компьютерного зрения с использованием глубокого обучения для управления подконтрольными системами. Результаты исследования основаны на использовании сверточной нейронной сети и направлены на решение задач как напрямую, так и косвенно связанных с проблемами в области экологии. Описан принцип и показана базовая модель использования технологии компьютерного зрения в актуальных проектах: 1) соблюдение нормы социально-дистанцирования в условиях пандемии COVID; 2) бесконтактный манипулятор для управления техническими объектами. Проект «Социальное дистанцирование в условиях пандемии COVID-19» разработан как средство автоматического контроля за дистанцией в местах массового скопления людей. Изучается принцип работы детектора, основанный на технологии компьютерного зрения, что позволяет контролировать соблюдение нормы физического дистанцирования. Для работы детектора требуется компьютер с необходимым

Образец цитирования:

Ткаченко СВ, Смирнова ТВ. Система компьютерного зрения в мобильных устройствах. *Журнал Белорусского государственного университета. Экология*. 2022;4:4–14.
<https://doi.org/10.46646/2521-683X/2022-4-4-14>

For citation:

Tkachenko SV, Smirnova TV. Computer vision system in mobile devices. *Journal of the Belarusian State University. Ecology*. 2022;4:4–14. Russian.
<https://doi.org/10.46646/2521-683X/2022-4-4-14>

Авторы:

Сергей Витальевич Ткаченко – аспирант кафедры информационных технологий в экологии и медицине.
Татьяна Васильевна Смирнова – кандидат физико-математических наук, доцент; доцент кафедры информационных технологий в экологии и медицине.

Authors:

Sergei V. Tkachenko, postgraduate student at the department of information technologies in ecology and medicine.
freddy.clarck@yandex.ru
Tatsiana V. Smirnova, PhD (physics and mathematics), docent; associate professor at the department of information technologies in ecology and medicine.
smirnova@iseu.by

программным обеспечением и результаты изображений с видеокамеры. Работа детектора проверена в реальных условиях: результаты контроля удовлетворительные. Погрешность при определении расстояния не превышала 9 %. Предполагается, что такое устройство может использоваться различными службами, включая санитарные, для контроля соблюдения населением рекомендаций системы здравоохранения относительно правил дистанцирования. Проект «Бесконтактный манипулятор» был разработан для дистанционного управления объектами при работе в труднодоступных и опасных для здоровья человека местах (атомные станции, переработка токсичных отходов), а также людям с ограниченными физическими возможностями. В его работе реализованы три функции – трекинг ключевых точек руки, перехват действий манипулятора компьютерной мыши, управление дополнительными функциями компьютера. Работа манипулятора также проверена в реальных условиях, степень достоверности результатов высокая. В перспективе манипулятор предполагается использовать в мобильном аппаратно-программном комплексе, предназначенному для обработки изображений с последующим анализом данных. Области практического применения комплекса – экологический мониторинг, исследование процессов эрозии почв, наблюдение за изменениями растительного покрова, анализ заболеваемости растений, сортировка отходов, экология человека. Ресурсы современных информационных технологий, реализованные через представленные проекты, демонстрируют возможность замены производства части материальных приборов их виртуальными аналогами, безопасными с точки зрения экологии.

Ключевые слова: компьютерное зрение; искусственная нейронная сеть; сверточная нейронная сеть; анализ изображений; L1-регуляризация; детектор объектов YOLO; NVIDIA CUDA; ключевые точки руки; фреймворк MEDIAPIPE; дополненная реальность.

COMPUTER VISION SYSTEM IN MOBILE DEVICES

S. V. TKACHENKO^a, T. V. SMIRNOVA^a

^a*International Sakharov Environmental Institute, Belarusian State University,
23/1 Daūhabrodskaja Street, Minsk 220070, Belarus
Corresponding author: S. V. Tkachenko (freddy.clarck@yandex.ru)*

The article presents the results of designing a computer vision system using deep learning to control supervised systems. The results are based on the use of convolutional neural network and aimed at solving problems, both directly and indirectly related to the solution of environmental problems. The principle is described and the basic model of using computer vision technology in actual projects is shown: 1) social distancing norm compliance in COVID pandemic; 2) non-contact manipulator to control technical objects. The COVID-19 Social Distancing Project was developed as a means of automatic control of social distance in crowded places. The article describes the principle of operation of the detector, based on computer vision technology, and makes it possible to monitor compliance with the norm of physical distance. To work the detector requires a computer with the necessary software and the results of images from a video camera. The operation of the detector has been tested under real conditions; the test results are satisfactory. The error in determining the distance did not exceed 9 %. It is assumed that such a device can be used by various services, including sanitary services, to monitor the compliance of the population with the recommendations of the health care system regarding the rules of distancing. The «Contactless Manipulator» project was designed for remote control of objects when working in hard-to-reach and hazardous places (nuclear power plants, toxic waste processing), as well as for people with limited physical abilities. Three functions of “Noncontact manipulator” were implemented in the work – tracking of key points of hand, interception of computer mouse manipulator actions, and control of additional computer functions. The work of the manipulator was also tested in real conditions, the degree of reliability of the results is high. In the future, the manipulator is supposed to be used in a mobile hardware-software complex designed for image processing with subsequent data analysis. The expected fields of practical application of the complex are ecological monitoring, investigation of soil erosion processes, observation of vegetation changes, analysis of plant diseases, waste sorting, and human ecology. The resources of modern information technologies, implemented through the presented projects, demonstrate the possibility of replacing the production of some material devices with their virtual counterparts, safe in terms of ecology.

Keywords: computer vision; artificial neural network; convolutional neural network; image analysis; L1-regularization; YOLO object detector; NVIDIA CUDA; key hand points; MEDIAPIPE framework; augmented reality.

Введение

Компьютерное зрение и искусственные нейронные сети – наиболее востребованные направления современных научных и прикладных исследований. Программно-аппаратные комплексы с использованием технологий компьютерного зрения способны решать задачи распознавания объектов

на основе наборов данных (например, в экологии); усилить возможности диагностики и проведения лечения (в медицине); сократить время на обслуживание клиентов (в логистических центрах); много-кратно увеличить возможности систем контроля и безопасности. Создание в настоящее время автоматизированных систем и аппаратно-программных комплексов на базе систем компьютерного зрения – одна из приоритетных технических задач. Компьютерное зрение – область искусственного интеллекта, связанная с анализом изображений и видео. Она включает в себя набор методов, которые позволяют на основе заданных параметром автоматически фиксировать и обрабатывать изображения неподвижных и движущихся объектов с помощью программно-аппаратных средств [1].

Мощности современных вычислительных устройств, повышение разрешающей способности камер видеонаблюдения, развитие полосы пропускания каналов связи, а также внедрение технологий глубокого обучения – все это способствует тому, что компьютерное зрение стало находить все больше применений в различных сферах жизнедеятельности. Основной рост воплощения решений с использованием компьютерного зрения произошел относительно недавно, когда стали активно использоваться искусственные нейронные сети при анализе и обработке данных. В простых системах обработки информации, основанных на компьютерном зрении, обычно требуется получить количественную и качественную информацию из визуальных данных. В основном это такие параметры, как размер, цвет, количество, направление и характер движения объекта, а также контрастные переходы в окрестностях пикселя изображения, из которых можно выделить характерные особенности. На основе этих параметров производится анализ изображения для извлечения нужной информации.

Компьютерное зрение и глубокое обучения сегодня широко используются при решении задач как непосредственно, так и косвенно связанных с экологическими проблемами: анализ аэрофотоснимков местности; вырубка лесов, браконьерство, пожары; незаконное использование земель под сельскохозяйственные угодья; промышленное и гражданское строительство; загрязнение лесных массивов промышленными (включая радионуклиды) и бытовыми отходами.

Актуальность работы обусловлена тем, что востребованы «автономные» аппаратно-программные комплексы, включающие компонент компьютерного зрения, со следующими условиями: относительно невысокая стоимость, доступная элементная база, собственное программное обеспечение. Такие устройства могут быть использованы различными службами, включая экологические.

Цель работы: представить результаты применения технологии компьютерного зрения при реализации проектов по разработке детектора социального дистанцирования, востребованного в период эпидемий и сезонных заболеваний, и манипулятора, заменяющего прибор для управления объектом его виртуальным аналогом.

Действующие аналоги системы социального дистанцирования имеются в США, Китае и Сингапуре [2]. В США используется система распознавания объектов *YOLO* для идентификации людей без масок. В Китае стартапы *SenseTime*, *Megvii* и *DeepGlint* предложили похожие технологии, которые, помимо контроля дистанции, также измеряют температуру людей в толпе. Алгоритмы анализируют цвет белков глаз, кожи и общее положение тела для определения вероятности заражения COVID-19. В проекте *AI Hub* (Сингапур) разработано приложение *SafeDistancer*. Оно реализует функцию преобразования смартфона в устройство для определения расстояния и контроля социальной дистанции. Все указанные разработки проходят многоэтапное тестирование, однако реализованы лишь единичные экземпляры.

Материалы и методы исследования

Усовершенствование структуры искусственных нейронных сетей. При использовании искусственной нейронной сети (ИНС) решаются вопросы ее обучения или самоорганизации. Для обучения нейронной сети необходимо иметь большой объем входной информации, поскольку добиться высокой точности работы ИНС можно только на большом наборе данных. Например, первая представленная версия распознавания образов на основе ИНС содержала 38-процентный уровень ошибок классификации, а сейчас ошибка не превышает 3 %. Первые результаты распознавания голоса на основе ИНС содержали 27 % ошибок распознавания, а сегодня процент ошибок составляет не более 8 % [3]. Низкий процент ошибок объясняется тем, что в настоящее время в качестве обучающих выборок используется большой объем доступной информации с видео-хостинга и серия аудиоклипов с описаниями.

Представленные в статье проекты основаны на использовании сверточных нейронных сетей. Они являются одними из наиболее популярных в глубоком виде машинного обучения, при котором модель учится выполнять задачи классификации непосредственно на изображении, видео, тексте. Перед сверточными нейронными сетями ставятся такие задачи классификации, как идентификация объекта,

определение границ и семантическая сегментация, выделение объектов внимания на изображении, распознавание лиц и частей тела, выделение нормалей к поверхности для реконструкции плоских изображений в трехмерное представление. Задача обучения ИНС состоит в настройке весов w нейронов для нахождения состояния, минимизирующую целевую функцию на обучающей и тестирующей выборках. Поиск глобального минимума функции реализуется методом градиентного спуска по функции потерь $F(w)$. Обновление весов ИНС происходит на основе итерационной процедуры обратного распространения ошибки.

Наиболее часто нейронная сеть строится в виде связного графа, особенностью которого является прямое распространение информации, и обратное распространение сигнала об ошибке. Один из недостатков такой топологии – избыточность. Как следствие, аппаратно-программная реализация такого алгоритма сети требует больших вычислительных мощностей. Возможное решение проблемы состоит в том, чтобы из всего набора факторов отобрать лишь часть важнейших и убрать малозначимые веса, производя фактически сжатие сети (рис. 1).

Обнуление весов позволяет улучшить обобщающую способность сети. Кроме того, сеть с большим количеством нулей занимает меньше места и способна быстрее рассчитаться на более слабых архитектурах, например, одноплатных компьютерах [4].

Достичь разреженности позволяют L -регуляризации (набор методов), которые могут повысить точность модели, представленной ИНС, за счет подавления незначащих факторов.

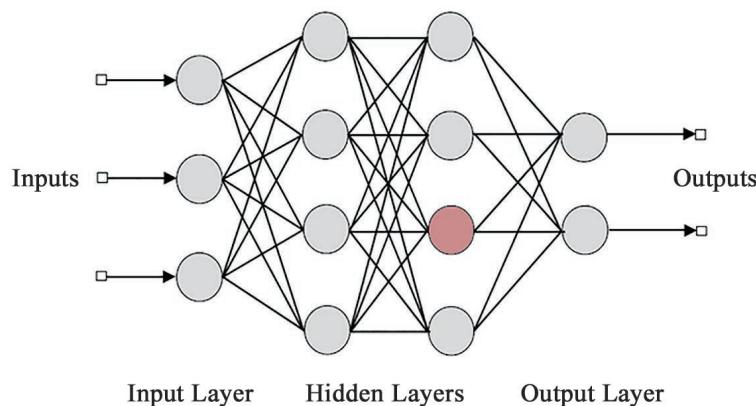


Рис. 1. Выявление «слабого» нейрона в искусственной нейронной сети

Fig. 1. Identifying the «weak» neuron in an artificial neural network

Существует три основных метода регуляризации: $L1$, $L2$ и $dropout$ [5]. $L1$ - (*lasso*) и $L2$ -регуляризации, которые подавляют подгоночные коэффициенты и квадраты весов, соответственно. В $dropout$ -регуляризации определенные нейроны исключаются из сети еще на этапе обучения. В результате приходим к более простой версии ИНС.

Однако эффект сжатия сети и ускорения предсказаний будет более значительным, если исключать из рассмотрения не отдельные веса, а нейроны из полностью связанных слоев. Наиболее простой и эффективный способ изъятия лишних нейронов из сети состоит в использовании групповой *lasso*-регуляризации. Ее применяют, чтобы держать малозначимые веса в сетях нулевыми. Суть метода состоит в применении специального маскирующего слоя с собственным вектором весов $M = (\beta_1, \beta_2, \dots, \beta_n)$. Вес маски накладывает ограничения на веса слоев, связанных своими выводами с маской. Если среди этих весов половина полезных, то значение β_i близко к единице, и этот вывод сможет хорошо передавать информацию. Если β_i близко к нулю, то это обнулит вывод нейрона и, заодно, все веса, от которых зависит этот вывод. После отбрасывания каналов проводится доучивание сети.

К недостаткам $L1$ -регуляризации относится то, что отбрасывание каналов может внести изменения в топологию и исказить первоначальную структуру сети на этапе обучения. Для ее решения применяют периодическое исключение нейронов. Подбор параметров маскирующего слоя реализован в методе вариационной оптимизации ($L0$ -регуляризации).

На рис. 2 представлены результаты сравнения $L1$ и $L0$ алгоритмов редукции каналов после серии экспериментов с разными константами мощности регуляризации. ИНС содержит 4 сверточных и 2 полностью связанных слоя; использовался набор данных *CIFAR-10* [6]. Сеть проходила обучение 200 эпох, после чего сохранялась, и к ней применялись алгоритмы редукции нейронов.

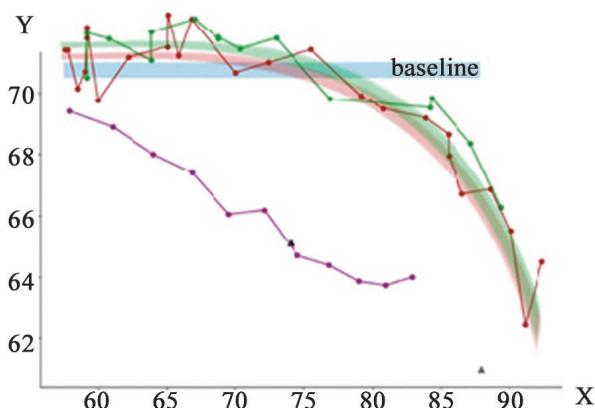


Рис. 2. Результат групповой регуляризации

Fig. 2. The result of group regularization

Здесь по оси X отложено уменьшение количества весов (в %) после применения соответствующего алгоритма редукции нейронов; по оси Y – точность сети с отброшенными нейронами на контрольной выборке. Синяя полоса – примерное качество сети, еще не подвергнутой редукции нейронов; зеленая линия – простой алгоритм $L1$ -обучения; красная линия – $L0$ -регуляризация; фиолетовая линия – удаление первых k каналов. Черные треугольники – обучение сети, у которой изначально было меньшее количество весов.

Вывод по обучению сети на доступном наборе данных следующий: многие каналы в обученной сети избыточны. Сокращение весов достигается с помощью регуляризации. На большом наборе данных классический $L1$ -алгоритм показывает результаты обучения сети, сопоставимые с $L0$ -регуляризацией.

Результаты исследования и их обсуждение

Применение технологии компьютерного зрения в детекторе социального дистанцирования. Соблюдение социального дистанцирования важно для снижения вероятности контакта между больными и здоровыми людьми. Автоматизированный контроль за соблюдением нормы социального дистанцирования особенно актуален в период инфекционной активности. Наличие достаточного количества средств контроля за соблюдением рекомендуемой нормы в 1,5–2 м позволяет дисциплинировать население и сдерживать распространение опасных инфекций.

Аппаратная составляющая детектора дистанцирования. Обнаружение объектов – технология, которая относится к области компьютерного зрения. Задача состоит в распознавании определенных объектов на статичном оцифрованном изображении или в видеопотоке. При этом каждый объект обладает набором характерных особенностей, с помощью которых его можно классифицировать. Методы обнаружения объектов основаны на глубоком обучении и использовании искусственных нейронных сетей. Существующие системы по выявлению необходимых объектов в режиме реального времени требуют или специального оборудования, включая мощные компоненты для обработки логических и графических операций, или наличия высокоскоростного Интернета для обработки цифровых данных на серверах. В качестве необходимого элемента системы используются и оптические компоненты – камеры видеонаблюдения.

Обработка дискретных цифровых изображений требует определить геометрические параметры объектов – расстояние, угол наклона, координатные преобразования. Главной проблемой при создании аппаратной части детектора по соблюдению дистанции является определения точного расстояния между контролируемыми объектами, в данном случае – дистанции между людьми. Точнее, нужно определить, соблюдается ли расстояние между объектами более чем в N пикселей (точек на прямоугольной сетке). Следовательно, для использования данных оптических приборов в системе необходимо выяснить, какое количество пикселей (двумерных наборов входных данных) будет оптимальным. Для этого предлагается использовать метод подобия.

Пусть есть маркер или объект с известной шириной W . Он помещается на некотором расстоянии D от камеры. До данным из камеры оцениваем ширину P в пикселях, что позволяет нам рассчитать фокусное расстояние F [7]:

$$F = (P \cdot D) / W. \quad (1)$$

Перемещая камеру к объекту/маркеру и применяя свойство подобия треугольников, можно определить расстояние от объекта до камеры по формуле:

$$D' = (W \cdot F) / P. \quad (2)$$

Полученные данные сохраняются в виде 2D-массивов точек и поступают в качестве входных данных программе по обработке.

При выполнении работы использовалась веб-камера *Logitech HD Pro Webcam C920* с такими характеристиками: количество точек матрицы 3Мп; автоматическая фокусировка Carl Zeiss Tessar, 20 шагов; максимальное разрешение видео 1920'1080; максимальное количество кадров/сек 30.

Программная составляющая детектора социального дистанцирования. Кроме аппаратной части системы, для ее управления требуется программное обеспечение. Программная составляющая детектора социального дистанцирования основывается на функционале библиотеки OpenCV для работы с изображениями, глубоком обучении, детекторе объектов YOLO, программно-аппаратной архитектуре параллельных вычислений NVIDIA CUDA GPU. Глубокое обучение – область машинного обучения, которая рассматривает методы решения задач искусственного интеллекта с использованием глубоких нейронных сетей. Основная проблема применения глубоких нейросетей с большим количеством скрытых слоев – трудность обучения по причине исчезающего градиента: при использовании традиционных функций активации сигналы об обратно распространяемых ошибках быстро становятся или очень малыми, или большими.

Для методов, основанных на глубоком обучении, используют главным образом архитектуру сверточных нейронных сетей (СНС) [8]. Этот тип ИНС обеспечивает устойчивость к изменению масштаба, повороту, смещению, смене ракурса, тем самым хорошо подходят для работы как со статическими, так и с динамическими изображениями (рис. 3).

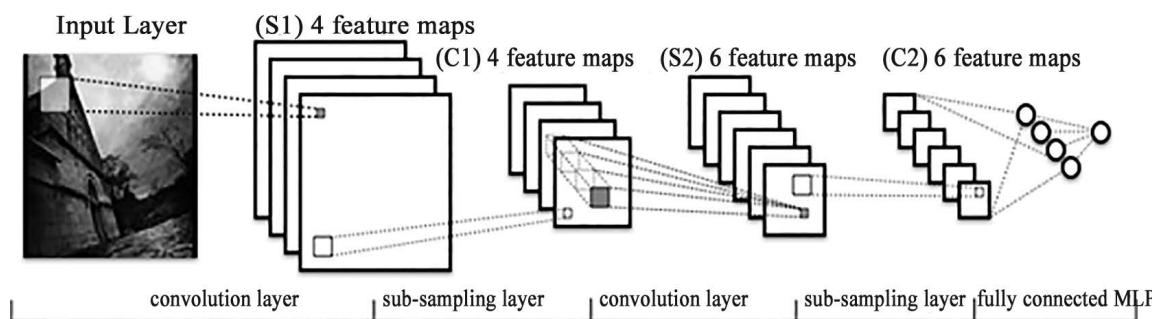


Рис. 3. Сверточная нейронная сеть

Fig. 3. The convolutional neural network

СНС позволяет обнаруживать объект без использования набора его специфических особенностей. Такая сеть обычно интерпретируется как переход от конкретных особенностей изображения к более абстрактным деталям. При этом сеть самонастраивается и вырабатывает необходимую иерархию абстрактных признаков, фильтруя маловажные детали и выделяя существенные. Основная идея состоит в чередовании слоев и применении операции свертки. Первые два типа слоев (сверточные и субдискретизирующие), чередуясь между собой, формируют входной вектор признаков для «обычной» сети – персептрона. Для ускорения обучения на больших наборах данных рекомендуется распараллеливать вычисления на многоядерных процессорах и на графических процессорах GPU. NVIDIA CUDA GPU (Compute Unified Device Architecture) является программно-аппаратной архитектурой параллельных вычислений, которая позволяет существенно увеличить вычислительную производительность.

Для решения задачи обнаружения объектов использовалась сверточная нейронная сеть *YOLOv4*. По литературным данным [8], эта версия нейронных сетей более быстрая и точная для решения задач идентификации объектов в режиме реального времени; способна распознавать несколько объектов в одном кадре; может указывать граничную рамку вокруг объекта. Достигается это тем, что *YOLO* «прогоняет» изображение через СНС один раз, в отличие, например, от двухэтапных методов [9]. Изображение делится на сетку. Каждой ячейке сетки принадлежит область объекта на изображении, если центр области находится в пределах ячейки. Для каждой области определяется класс объекта и охватывающая рамка. Что касается оценки эффективности работы ИНС (показатели точности и скорости), то отметим следующее. Поскольку в данной статье представлены результаты исследовательской работы, а отладка проводилась с использованием вычислительных ресурсов CPU, можно говорить лишь о значительном увеличении производительности архитектуры *YOLO* при использовании GPU, включая распараллеливание алгоритмов. Обоснованные цифры будут приведены в следующей работе. С учетом этого разработана программа на языке Python, способная выполнять анализ как по стационарным видеоданным, так и в режиме реального времени.

Поскольку детектор дистанцирования разрабатывался под разные архитектуры, то в программной части предусмотрен ввод значений двух констант:

- логическое значение – указывающее (предполагается использование CPU или GPU в зависимости от этого корректируются параметры конфигурационного файла);
- минимальное расстояние (в пикселях) для соблюдения социальной дистанции.

Кроме того, предусмотрена обработка изображения как в реальном времени (видеопоток данных), так и в режиме симуляции (с файлом данных, подготовленным заранее). При использовании готового файла системой предусмотрена возможность подготовки измененного видеофайла с метками на анализируемых объектах. При разработке программы использовался алгоритм обработки из библиотеки TensorFlow [6]. Внесенные в алгоритм изменения позволяют сократить объем обрабатываемой информации за счет исключения факторов с малыми весами и реализовать дополнительный функционал.

Обучение сети в задаче детектирования проводилась на наборе данных CIFAR10. Набор данных имеет точность около 99,37 %, при обучении в 200 эпох. Разработанная же нами модель при обучении проходила 100 эпох, при этом точность составила 82 %.

Показатель точности выводился в программе; фрагмент кода приведен ниже:

```
score = model.evaluate(X_test, y_test,  
                      batch_size=32, verbose=1)  
print()  
print(u'Оценка теста: {}'.format(score[0]))  
print(u'Оценка точности модели: {}'.format(score[1]))
```

Для тестирования и получения результатов использовались кадры видеоизображений. Отобранный вручную набор был разбит на 3 класса в зависимости от наличия различных помех, условий освещенности, числа людей в кадре. Точность для всех классов составила в целом величину 79,7 %.

Программа позволяет получить следующие результаты:

- обнаруживать наличие объекта в видеопотоке;
- выделять ограничивающую рамку;
- определять центроиды для каждого обнаруженного человека;
- производить вычисление попарных расстояний между всеми центроидами;
- производить проверку, выдерживается ли рекомендуемое расстояние между объектами. Если обнаруживается, что парные расстояния менее N пикселей, то это означает, что пара людей нарушила правила социального дистанцирования (рис. 4).



Рис. 4. Применение детектора социального дистанцирования на практике:
слева – объекты, нарушающие дистанцию; справа – прошедшие обработку детектором

Fig. 4. Application of the social distancing detector in practice:
left – objects that violate the distance; right – objects that have been treated by the detector

Тестирование программы проводилось на CPU, но в коде предусмотрена возможность использовать графический процессор NVIDIA CUDA, который поддерживает режим реального времени, что позволяет использовать детектор в качестве мобильного прибора по проверке соблюдения норм дистанцирования.

Проектирование системы компьютерного зрения для альтернативного управления устройства в виртуальной среде жестами руки. При создании автоматизированных систем с применением компьютерного зрения в первую очередь предполагается получать результаты обнаружения объектов при обработке видеоданных. Не менее важным элементом является организация естественного интерфейса «человек – техническое устройство», основанного на жестах, мимике, голосовых командах. Кроме того, возможна реализация зрительной обратной связи при работе управляемых устройств и созданию механических манипуляторов. Это особенно востребовано в робототехнике, в индустрии дополненной и виртуальной реальности, в обучающих системах.

Дополненная реальность – это среда, дополняющая реальный физический мир цифровыми данными с помощью каких-либо устройств. В такой среде виртуальные объекты проецируются на реальное окружение, а взаимодействие между реальными и виртуальными объектами создается с помощью компьютерного зрения.

В рассматриваемой части работы разработан и реализован алгоритм по созданию и функционированию компонента дополненной реальности, выполняющий функции регулятора звука и нажатия кнопок мыши.

Обнаружение ключевых признаков руки с использованием машинного обучения. Обнаружение ключевых признаков (точек) руки – это процесс нахождения суставов и кончиков пальцев на заданном изображении [10]. Для поиска ключевых точек предлагается использовать фреймворк *MediaPipe* с открытым исходным кодом [11], который помогает создавать мультимодальные конвейеры машинного обучения.

Структура *MediaPipe* состоит из трех основных компонентов – основа для вывода из данных конвейера, инструменты для оценки и набора повторно используемых компонентов логического вывода и обработки. Фактически *MediaPipe* использует модель однократного обнаружения руки (ладони). И как только границы ладони будут обнаружены, фреймворк выполняет точную локализацию ключевых точек в пределах границ и выдает координаты 21 точки. Таким образом, происходит захват двадцать одной точки выпуклых частей кисти (рис. 5).

Ключевые точки позволяют восстановить полную модель кисти по ее входному изображению. Это наиболее перспективный визуальный метод распознавания жестов, так как он подразумевают полное оценивание текущего положения и динамики руки. Но метод достаточно сложен для программирования «с нуля» и ограничен по точности восстановления изображения.

В конвейере *MediaPipe* используется несколько моделей. Одна из них – модель обнаружения ладони, которая возвращает ограничивающую и ориентированную в пространстве рамку руки на основе ее полного изображения [11].

Модель и реализованный для нее метод были взяты за основу и адаптирован в виде программного модуля на языке *Python* [12]. Входное изображение – реальная ладонь разработчика. Последовательные шаги реализации алгоритма следующие:

- 1) детектор определяет пространственную скелетную модель изображения руки;
- 2) с помощью ограничивающей рамки выделяется область, содержащая полезную информацию;
- 3) выделенная область подается на вход модели, определяющей ориентир руки. Результат – набор трехмерных координат скелета руки (рис. 5);
- 4) распознаватель жестов отыскивает конфигурацию 3D-координат в наборе жестов.

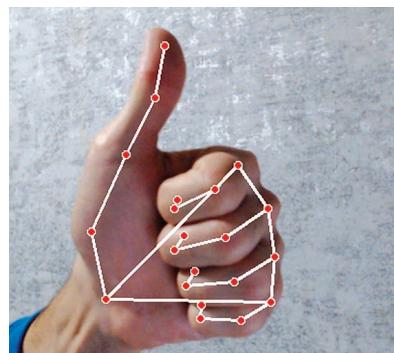


Рис. 5. Ключевые точки в 3D-измерении

Fig. 5. 3D key points

Правильно определенная форма ладони позволяет значительно ускорить решение задачи распознавания жестов, что приводит к сокращению размерности данных и возможности использовать такой алгоритм на одноплатных компьютерах.

При тестировании программного модуля были проведены эксперименты по регулировке звука и нажатию клавиш компьютерной мыши.

Управление звуком при помощи жестов. Модель руки создает локализацию руки на основе указанных ключевых точек. Результат локализации оформляется в виде программного модуля. При обращении к модулю на вход подается реальное изображение ладони. После обработки изображения будет получен массив данных (вектор) v , содержащий 3D-координаты 21 точки ладони.

Используя массив данных с координатами точек руки, на языке *Python* был реализован код, осуществляющий управление звуком жестами руки. Использовались элементы векторов указанных ключевых точек (рис. 6).

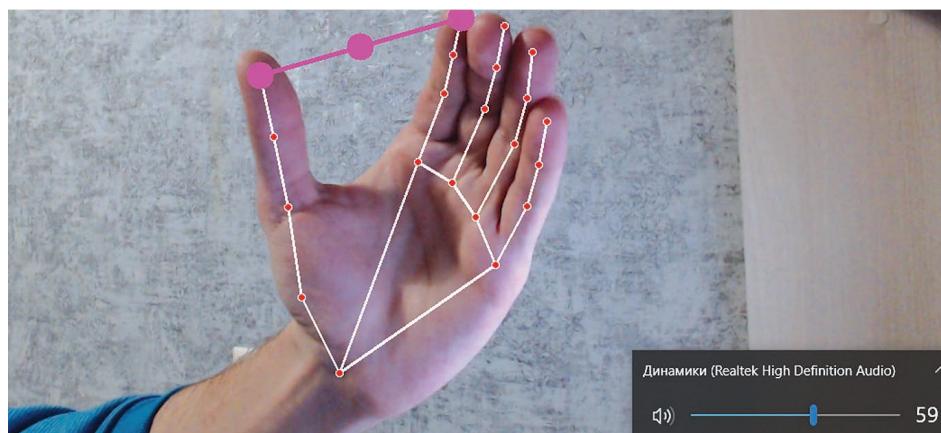


Рис. 6. Регулирование звука жестами

Fig. 6. Gesture-based audio control

При тестировании программного модуля замечено, что точность и скорость регулировки сильно зависят от расстояния между управляемым объектом и камерой.

По поводу определения расстояния стоит заметить, что на сегодняшний день нет алгоритма точного расчета такого расстояния [12]. Более того, нейронные сети принадлежат к непараметрическим моделям, поэтому мера сходства в виде функции расстояния здесь не используется.

Результаты проведенных экспериментов показали, что наилучшее расстояние от камеры $l \sim 500$ м. При уменьшении расстояния действий по регулировке требовалось меньше, но и точность падала. При увеличении расстояния произвести регулировку не удавалось. Таким образом, в видеопотоке можно определить не абсолютный, а только относительный размер и расстояние между объектами.

Манипулирование двухкоординатной мышкой при помощи жестов. Ограничения обычной компьютерной мышки проявляются в том, что при 3D-проектировании уже недостаточно возможности манипуляций с объектом на плоскости. В качестве варианта решения такой проблемы предлагается управлять движением объекта с помощью жестов руки. Жесты позволяют организовать сложные типы движений, включая повороты. Кроме того, такой манипулятор может управлять удаленным объектом.

Реализация алгоритма по управлению движением жестами включает следующие шаги:

- получение выделяющей рамки руки и координат ключевых точек;
- соотношение координат поверхности с мышью и рамки руки;
- управление перемещением курсора и нажатием кнопок.

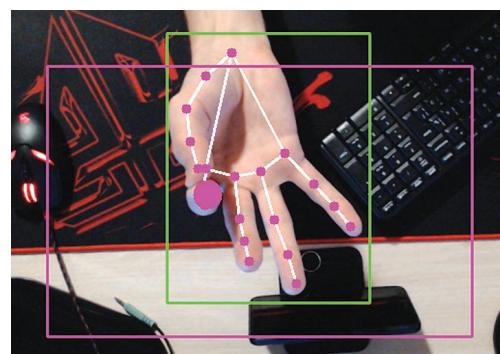


Рис. 7. Регулировка нажатий кнопок мышки

Fig. 7. Adjusting the mouse buttons

Как и в системе управления звуком, основная трудность при реализации состояла в корректном определении расстояния от камеры. Множеством экспериментов по определению расстояния было установлено, что, как и в предыдущем случае, расстояние от камеры $l \sim 500$ мм позволяло за минимальное число управляющих жестов добиться требуемого результата (рис. 7). С другой стороны, это позволит в дальнейшем комбинировать две разработанные модели в общую систему.

Заключение

Система компьютерного зрения как отрасль знания ориентирована прежде всего на практическое применение. Алгоритмы обработки изображений трудоемкие, их реализация требует большого объема расчетов. Еще более мощные вычисления требуются при обработке изображений в режиме реального времени. Учитывая, что основные приложения машинного зрения принадлежат к областям, где характеристики конструируемых устройств и по объему, и по стоимости должны быть небольшими, особый интерес представляет программная реализация для маломощных вычислительных устройств и аппаратно-программных комплексов.

В статье представлены результаты разработки макетов устройства, основанные на технологии компьютерного зрения с использованием СНС. Устройство предназначено для контроля соблюдения норм социального дистанцирования и может использоваться соответствующими службами в местах массового обслуживания людей.

Как перспективное и безопасное в области экологии направление практического применения технологии компьютерного зрения – альтернативное управление техническими устройствами с помощью команд, подаваемых жестами, голосом. В работе представлен и реализован алгоритм по созданию интерфейса «человек – компьютер», основанный на распознаваемых компьютером жестах и исключающий «удобные» для компьютера клавиатуру и мышь. Преимущество такого типа управлений в том, что они работают в реальном времени, без специального оборудования, управляют устройствами, не касаясь их физически.

Проведена серия экспериментов по подбору оптимального расстояния от управляемого объекта до камеры с целью получения минимального по объему, но достаточного для идентификации действий набора данных. Конечная задача – реализация алгоритма на одноплатных компьютерах. К настоящему времени не разработаны в достаточной степени методы экспресс-анализа и сбора данных об окружающей среде, основанные на исследовании фото-, видео- данных, и реализованные в виде автоматизированных систем. Научная новизна работы состоит в совершенствовании методов и средств для автоматизации контроля за объектами окружающей среды на основе современных информационных технологий. Технологии компьютерного зрения – универсальный и мощный инструмент исследования, который можно применять в различных областях, в первую очередь к проблемам экологии. Однако сегодня реализация данной технологии в проектах и внедрение их в повседневную жизнь затруднена по ряду причин (недостаточная элементная база, финансовые риски и технофобия).

Библиографические ссылки

1. Bishop CM. Pattern Recognition and Machine Learning. New York: Springer-Verlag; 2006. 738 р.
2. Sense Time: Artificial Intelligence (AI) Software Provider [Интернет; процитировано 25 августа 2022]. URL: www.sensetime.com/en.
3. Ткаченко СВ, Смирнова ТВ. Совершенствование структуры искусственных нейронных сетей на базе вариационной оптимизации. В: Сахаровские чтения 2020: экологические проблемы XXI века». Часть 2. с. 322–326. DOI:10.46646/SAKH-2020-2-322-326
4. Розенблatt Ф. Принципы нейродинамики. Перцептроны и теория механизмов мозга. Москва: Мир; 1965. 302 с.
5. Duchi, J., Hazan E., Singer Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*. 2011;5:2121–2159.
6. Popular Datasets Over Time [Интернет; процитировано 25 марта 2022]. URL: www.kaggle.com/code.
7. Бринк Х, Ричардс Дж, Феверолф М. Машинное обучение. Санкт-Петербург: Питер; 2018. 336 с.
8. Wang Ch, Bochkovskiy A., Liao M. Scaled-YOLOv4: Scaling Cross Stage Partial Network [Интернет; процитировано 25 августа 2022]. URL: <https://arxiv.org/abs/2011.08036v2>.
9. Boominathan L, Kruthiventi S, Babu R. CrowdNet: A Deep Convolutional Network for Dense Crowd Counting. *Proceedings of the 2016 ACM on Multimedia Conference*. Amsterdam, Netherlands. 2016: 640–644. DOI:10.1145/2964284.2967300
10. Cao Z, Xu X, Biao Z, Zhou M, Li Q. Real-time gesture recognition based on feature recalibration network with multi-scale information. *Neurocomputing*. 2019;347:119–130.
11. MediaPipe Python Framework [Интернет; процитировано 12 февраля 2022]. URL: <https://google.github.io/mediapipe/solutions/solutions.html>.
12. Шакирьянин ЭД. Компьютерное зрение на Python. Москва: Лаборатория знаний; 2021. 163 с.

References

1. Bishop CM. *Pattern Recognition and Machine Learning*. New York: Springer-Verlag; 2006. 738 p.
2. Sense Time: Artificial Intelligence (AI) Software Provider [Internet; cited 2022 August 25]. Available from: www.sensetime.com.
3. Tkachenko SV, Smirnova TV. *Sovershenstvovaniye struktury iskusstvennykh nevronnykh setey na baze variatsionnoy optimizatsii* [Improving of the structure of artificial neural networks based on variational optimization]. In: *Sakharovskie chteniya 2020: ekologicheskie problemy XXI veka. Part 2*. p. 322–326. DOI:10.46646/SAKH-2020-2-322-326. Russian.
4. Rosenblatt F. *Printsipy neyrodinamiki. Pertseptrony i teoriya mekhanizmov mozga* [Principles of neurodynamics. Perceptrons and the theory of brain mechanisms]. Moscow: Mir; 1965. 302 p. Russian.
5. Duch J, Hazan E, Singer Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*. 2011;5:2121–2159.
6. Popular Datasets Over Time [Internet; cited 2022 March 25]. Available from: www.kaggle.com/code.
7. Brink H, Richards J, Feverolf M. *Mashinnoye obuchenije* [Machine learning]. Saint Petersburg: Piter; 2018. 336 p. Russian.
8. Wang Ch, Bochkovskiy A., Liao M. *Scaled-YOLOv4: Scaling Cross Stage Partial Network* [Internet; cited 2022 August 25]. Available from: <https://arxiv.org/abs/2011.08036v2>.
9. Boominathan L, Kruthiventi S, Babu R. CrowdNet: A Deep Convolutional Network for Dense Crowd Counting. *Proceedings of the 2016 ACM on Multimedia Conference*. Amsterdam, Netherlands. 2016: 640–644. DOI:10.1145/2964284.2967300
10. Cao Z, Xu X, Biao Z, Zhou M, Li Q. Real-time gesture recognition based on feature recalibration network with multi-scale information. *Neurocomputing*. 2019;347:119–130.
11. MediaPipe Python Framework [Internet; cited 2022 February 12] Available from: <https://google.github.io/mediapipe/solutions/solutions.html>
12. Shakiryanov ED. *Komp'yuternoye zreniye na Python* [Computer vision in Python]. Moscow: Knowledge Lab; 2021. 163 p. Russian.

Статья поступила в редакцию 13.10.2022.

Received by editorial board 13.10.2022.