

**Conclusion.** We used a personal computer with an intel core i7 processor (2.21 GHz CPU) with the Ubuntu 18.04 operating system installed. When calculating the initial geometry of a molecule with an N-(2-hydroxy-3,5-diisopropylphenyl) benzene sulfonamide compound base, the method of molecular mechanics (MM+) of the Hyper Chem 08 software package was chosen. The maximum wavelength with a high oscillator strength was observed at  $\lambda = 291.63$  nm and  $f = 0.0441$  (Table, Fig. 2,3). The calculation showed that the strongest electron transition is observed at the absorption maximum of 291.63 nm, which refers to the electron transition to the excited singlet state  $S_0 \rightarrow S_8$ . The remaining transitions have a small value of  $f$  and are forbidden by symmetry.

#### REFERENCES

1. *Shahab, S.* Interaction between new synthesized derivative of (E,E)-azomethines and BN(6,6-7) nanotube for medical applications: Geometry optimization, molecular structure, spectroscopic (NMR, UV/Vis, excited state), FMO, MEP and HOMO-LUMO investigations / S. Shahab [at all] // Journal of Molecular Structure. – 2017. – Vol. 1. – No. 1146. – P. 881–888.
2. *Sheikhi, M., Shahab, S., Filippovich, L., Khaleghian, M., Dikisar, E., Mashayekhi, M.* Interaction between new synthesized derivative of (E,E)-azomethines and BN(6,6-7) nanotube for medical applications: Geometry optimization, molecular structure, spectroscopic (NMR, UV/Vis, excited state), FMO, MEP and HOMO-LUMO investigations // Journal of Molecular Structure. – 2017. – No. 1146. – P. 881–888.

## СИСТЕМА КОМПЬЮТЕРНОГО ЗРЕНИЯ И СОЗДАНИЕ ИНТЕРФЕЙСОВ УПРАВЛЕНИЯ В ДОПОЛНЕННОЙ РЕАЛЬНОСТИ

### COMPUTER VISION SYSTEM AND CREATION OF CONTROL INTERFACES IN AUGMENTED REALITY

**C. B. Ткаченко<sup>1,2</sup>, Т. В. Смирнова<sup>1,2</sup>**

**S. Tkachenko<sup>1,2</sup>, T. Smirnova<sup>1,2</sup>**

<sup>1</sup>Белорусский государственный университет, БГУ, г. Минск, Республика Беларусь

<sup>2</sup>Учреждение образования «Международный государственный экологический институт имени А. Д. Сахарова» Белорусского государственного университета, МГЭИ им. А. Д. Сахарова БГУ, г. Минск, Республика Беларусь

<sup>1</sup>Belarusian State University, BSU, Minsk, Republic of Belarus

<sup>2</sup>International Sakharov Environmental Institute of Belarusian State University, ISEI BSU, Minsk, Republic of Belarus  
[smirnova@iseu.by](mailto:smirnova@iseu.by)

Представлены реализованные алгоритмы по созданию интерфейса «человек-компьютер», основанные на технологии компьютерного зрения и решении задачи распознавания компьютером жестов пользователя.

Implemented algorithms for creating a human-computer interface based on computer vision technology and solving the problem of recognizing user gestures by a computer are presented.

*Ключевые слова:* компьютерное зрение, CV, анализ изображений, дополненная реальность, ключевые точки руки, фреймворк MEDIAPIPE.

*Keywords:* computer vision, CV, image analysis, augmented reality, hand keypoints, framework MEDIAPIPE.

<https://doi.org/10.46646/SAKH-2022-2-405-408>

**Введение.** Вычислительные мощности современной компьютерной техники, повышение качества и скорости обработки аудио- и видеoinформации способны во многом изменить качество жизни современного человека. Автоматизированные системы и аппаратно-программные комплексы, основанные на обработке информации с помощью технологий компьютерного зрения и искусственного интеллекта – наиболее приоритетные направления прикладных исследований.

Компьютерное зрение – это область искусственного интеллекта, связанная с фиксацией и обработкой неподвижных и движущихся объектов при помощи компьютера. С помощью зрения люди воспринимают окружающий мир. Но использование зрительной системы человека ограничено рядом объективных и субъективных особенностей, что требует использования дополнительных устройств. Компьютерное зрение включает в себя набор методов, наделяющих компьютер способностью «увидеть» и извлечь из увиденного полезную информацию. Для этого используют технологии машинного обучения.

Первые попытки заставить компьютер воспроизводить и автоматизировать зрительные функции человека относятся к середине 20 века. Но только в последние годы возможности компьютерного зрения стали находить все больше применений в различных отраслях. Основной рост внедрения решений произошел, когда стали более активно использоваться искусственные нейронные сети при поиске и обработке информации.

Современные решения с использованием компьютерного зрения находят применение в таких областях, как визуальный поиск объектов; распознавание текста и изображений; системы безопасности и контроля; навигация и транспортировка, робототехника, медицина и др.

Актуальность и практическая значимость работы обусловлена потребностью в автоматизированных устройствах, основанных на доступной элементной базе, и созданию для них программного обеспечения. Такие устройства, кроме оперативной обработки видеоданных, способны решать задачи организации 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.

**Системы и методы обработки изображений.** Системы компьютерного зрения (CV) состоят из фото- и/или видеокамер для фиксации изображения, и аппаратной части – компьютера, при помощи которого происходит анализ и обработка полученных изображений или видеофрагментов. Для анализа и обработки полученных фото- или видеоданных с использованием компьютерного зрения используют программные решения, собранные в различные библиотеки. Наиболее известные из них:

- *OpenCV* (Open Source Computer Vision Library) – библиотека алгоритмов компьютерного зрения, обработки изображений и численных алгоритмов общего назначения. Реализована на языке C/C++, также разрабатывается для Python, Java, Ruby, Matlab, Lua и других языках

- *Amazon Recognition* – платформа с возможностью глубокого анализа попавших в объектив камеры предметов, сооружений и людей. Сервис является частью системы с интегрированным механизмом полномасштабного самообучения. При использовании платформы для распознавания лица программное обеспечение определит по внешним признакам текущее эмоциональное состояние исследуемой личности. Методика включает в себя сравнение двух изображений на основе  $10^6$  признаков.

- *CUDA* (Compute Unified Device Architecture) – программно-аппаратная архитектура параллельных вычислений, которая позволяет существенно увеличить вычислительную производительность благодаря использованию графических процессоров фирмы Nvidia.

В простых системах обработки CV 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. Вся обработка в готовых программных решениях происходит на серверах, куда камеры передают данные. Передача данных через интернет замедляет или даже останавливает процесс обработки. Поэтому нужны «автономные» решения.

В области компьютерного зрения принято выделять следующие основные этапы обработки данных [1]:

1. предварительная обработка изображений, включая его преобразование в цифровую форму;
2. сегментация, с последующим анализом полученного результата;
3. выделение геометрической структуры;
4. идентификация или классификация изображения, его «понимание».

Уровни обработки, связанные с этими этапами, относятся к обработке нижнего, среднего, высокого уровня, соответственно. Наиболее исследованы алгоритмы обработки нижнего уровня (линейная и нелинейная фильтрация, гистограммная обработка). В разработке методов обработки среднего уровня (сегментация и выделение геометрической структуры) в последнее время достигнут значительный успех, особенно при выделении признаков простых геометрических структур внутри малых фрагментов («точка», «край», «прямая линия», «угол»). Разработка методов высокого уровня – перспективное направление, используется глубокое обучение и доступные датасеты.

**Использование системы компьютерного зрения как компонента дополненной реальности.** Дополненная реальность – это среда, дополняющая реальный физический мир цифровыми данными с помощью каких-либо устройств. В дополненной реальности виртуальные объекты проецируются на реальное окружение. Создается же взаимодействие между реальными и виртуальными объектами с помощью компьютерного зрения.

В практической части работы разработан и реализован алгоритм по созданию и функционированию компонента дополненной реальности, выполняющий функции регулятора звука и нажатий клавиш мышки.

**Обнаружение ключевых точек руки с использованием глубокого обучения и OpenCV.** Обнаружение ключевых точек руки – это процесс обнаружения суставов пальцев, а также кончиков пальцев на заданном изображении. Для поиска ключевых точек предлагается использовать фреймворк MediaPipe с открытым исходным кодом, представленный Google, 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.

По сути, 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. Другими словами, происходит захват двадцать одной точки выпуклых частей ладони руки (Рис. 1).



Рисунок 1 – Ключевые точки ладони

Ключевые точки позволяют восстановить полную модель ладони по ее входному изображению. Это наиболее перспективный визуальный метод распознавания жестов, так как он подразумевает 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. Метод был взят за основу и адаптирован в виде программного модуля на языке Python. В качестве входного изображения подавалась ладонь реального пользователя [2].

1. Детектор ладони определил пространственную скелетную модель изображения руки.
2. Далее с помощью ограничивающей рамки выделялась область, содержащая полезную информацию.
3. Выделенная область подавалась на вход модели, определяющей ориентир руки
4. В результате работы программного модуля были получены высокоточные трехмерные координаты ключевых точек руки (Рис. 2).

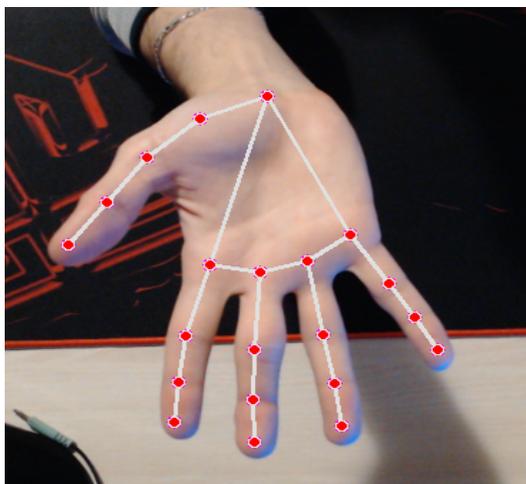


Рисунок 2 – Реальные ключевые точки в 3D-измерении

**Практическое применение технологии CV для управления устройствами.** Зная точные значения трехмерных координат каждой из 21 точки ладони, можно строить модели для разных практических применений. В частности, с помощью CV были реализованы интерфейсы между пользователями и компьютером, основанные на распознавании жестов и исключающие «удобные» для компьютера клавиатуру и мышь.

**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 руки. После обработки изображения в качестве результата будет получен массив данных (вектор), содержащий координаты 21 важной точки, образующих скелет руки. Точки всегда будут означать одно и то же, независимо от входного изображения. Например, точка №4 всегда будет кончиком большого пальца, а точка №8 всегда будет кончиком указательного пальца (Рис. 1, 3). Используя массив данных с координатами точек руки, на языке Python был реализован код, реализующий управление звуком с помощью жестов руки. Использовались элементы вектора  $v[4]$ ,  $v[5]$ ,  $v[8]$  (Рис. 3).

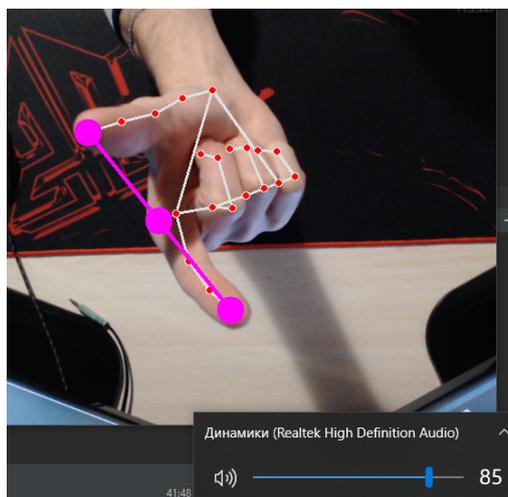


Рисунок 3 – Регулировка звука жестами

Следует отметить, что точность и скорость регулировки сильно зависят от расстояния между управляемым объектом и камерой. Точного алгоритма расчета такого расстояния нет, поскольку оно определяется во многом индивидуальными характеристиками. Расстояние определялось экспериментально,  $l=500$  мм от камеры. При уменьшении расстояния действий по регулировке требовалось меньше, но и точность падала. При увеличении расстояния произвести регулировку не удавалось.

Как вывод, можно утверждать, что в видеопотоке можно определить только относительный размер и расстояние между объектами. Для получения абсолютных размеров надо знать скорость и направление.

#### Управление нажатием кнопками мышки при помощи жестов.

Этапы по управлению компьютерной мышкой состоят в следующем (Рис. 4):

- поиск основных ключевых точек на руке и получение 3D-координат найденных точек;
- определение управляющих пальцев;
- соотношение координат рабочего стола и выделенной рамки руки;
- передвижение компьютерной мыши и обработка нажатий.

Как и в системе управления звуком, основные взаимодействия с пользователем с системой компьютерного зрения происходили на расстоянии  $l=500$  мм от камеры. Данное расстояние было отрегулировано на основании первой модели управления, что позволит в дальнейшем комбинировать две разработанные модели в одну систему.

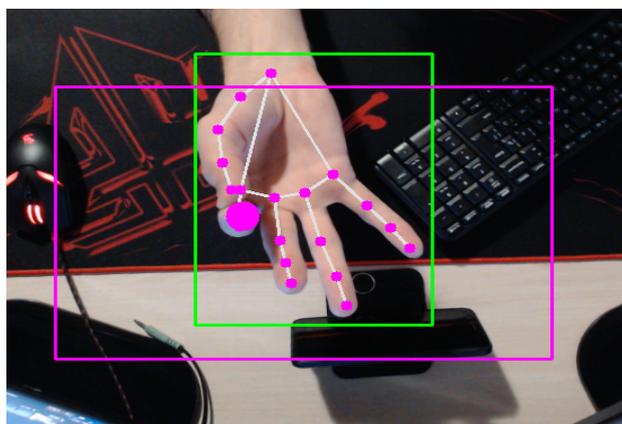


Рисунок 4 – Регулировка нажатий кнопок мышки

**Заключение.** Машинное зрение, как отрасль знания, ориентирована прежде всего на практическое применение. Алгоритмы обработки изображений трудоемки, а их реализация требует колоссального объема расчетов. Еще большие вычислительные затраты требуются при обработке изображений в режиме реального времени. Учитывая, что основные приложения машинного зрения принадлежат к областям, где характеристики конструируемых устройств как по объему, так и по стоимости должны быть небольшими, особый интерес представляет разработка алгоритмов для маломощных вычислительных устройств. В работе представлен и реализован алгоритм по созданию бесконтактного интерфейса «человек-компьютер», основанный на распознаваемых компьютером жестах.

#### ЛИТЕРАТУРА

1. Форсайт Д, Понс Ж. Компьютерное зрение. Современный подход. М.: Изд. дом «Вильямс», 2004 – 928 с.
2. Шакирьянов, Э.Д. Компьютерное зрение на Python. – Лаборатория знаний, 2021. – 160 с.