

**МРНТИ: 28.23.15**

Д.Ж. Сатыбалдина<sup>1</sup>, Н.С. Глазырина<sup>1</sup>, В.С. Степанов<sup>2</sup>, К.А. Калымова<sup>3</sup>

<sup>1</sup> *Евразийский национальный университет имени Л.Н. Гумилева, Нур-Султан, Казахстан*

<sup>2</sup> *ТОО «Kazdream Technologies», Нур-Султан, Казахстан*

<sup>3</sup> *Казахский национальный женский педагогический университет, Алматы, Казахстан  
(E-mail: satybalдина\_dzh@enu.kz, glazyrina\_ns\_1@enu.kz, rightcrow@gmail.com, gulzia\_kalymova@mail.ru)*

**Разработка Python приложения для распознавания жестов рук из видеопотока RGB и RGBD камер**

**Аннотация:** Системы распознавания жестов сильно изменились за последнее время, благодаря разработке современных устройств захвата данных (сенсоров) и разработке новых алгоритмов для распознавания. В статье представлены результаты исследования для распознавания статических и динамических жестов рук из видеопотока с RGB и RGBD камер, а именно с веб-камеры Logitech HD Pro Webcam C920 и с камеры глубины Intel RealSense D435. Программная реализация выполнена с использованием инструментов Python 3.6. Python-библиотеки с открытым исходным кодом обеспечивают надежную реализацию алгоритмов обработки и сегментации изображений. Подсистема извлечения признаков и классификации жестов построена на основе архитектуры нейронной сети VGG-16, реализованной с использованием фреймворков глубокого обучения TensorFlow и Keras. Приведены технические характеристики камер. Описан алгоритм работы приложения. Приведены исследования, направленные на сравнение устройств захвата данных, при различных условиях эксперимента (по расстоянию и освещенности). Экспериментальные результаты показывают, что использование камеры глубины Intel RealSense D435 обеспечивает более точное распознавание жестов при различных условиях эксперимента.

**Ключевые слова:** камера глубины, распознавание жестов, сверточная нейронная сеть, RealSense, OpenCV, Python, VGG-16.

DOI: <https://doi.org/10.32523/bulmathenu.2021/3.1>

**2000 Mathematics Subject Classification: 68T10**

**1. Введение.** В настоящее время многие исследования направлены на разработку и совершенствование методов человеко-машинного взаимодействия. В частности, большое внимание уделяется распознаванию жестов, поскольку его применение варьируется от систем распознавания языка жестов и преобразования жестов рук в речь [1-4] до медицинских систем реабилитации людей с ограниченными возможностями [5] и управления роботами [6].

Важным этапом в процессе распознавания жестов, является ввод жеста в систему с помощью различных сенсоров, от выбора которых зависит построение системы распознавания. Среди сенсоров для ввода жеста можно выделить сенсоры, основанные на вводе изображения в систему, к которым относятся видеокамеры, и не основанные на вводе изображения, примерами которых могут служить информационные перчатки и браслеты.

Второй тип сенсоров предоставляет системе точные данные о положении рук и конфигурации пальцев, однако недостатками их применения являются высокие требования

к калибровке, постоянное взаимодействие с генератором жестов и стоимость. Видеокамеры же не требуют постоянного физического контакта с пользователем и дают возможность пользователю совершать движения более естественным образом, кроме того развитие аппаратного и программного обеспечения видеокамер и реализация сопутствующих алгоритмов для распознавания жестов, делают их более удобными для применения в системах распознавания жестов [7].

В работах [8, 9] нами уже обоснован подход, выбор архитектуры нейронной сети, подготовлены наборы данных для обучения и тестирования нейронной сети и ее программная реализация для задачи распознавания статических жестов. Данные для нейронной сети были получены с камеры RGBD. Экспериментальные исследования показали высокую точность распознавания жестов (99,4%).

В настоящей статье мы представляем результаты исследований для распознавания статических и динамических жестов из видеопотока с веб-камеры Logitech HD Pro Webcam C920 и камеры глубины Intel RealSense D435, при различных условиях эксперимента по расстоянию до камер и освещенности. Выделение признаков и классификация жестов рук реализованы на основе предварительно обученной сверточной нейронной сети с архитектурой VGG-16. На основе полученных в режиме реального времени данных выявлены оценки производительности для двух видов камер, а также подобраны оптимальные условия эксперимента.

Остальная часть этой статьи структурирована следующим образом. В разделе 2 описаны связанные работы. В разделе 3 описан алгоритм работы приложения для распознавания жестов рук. В разделе 4 представлены экспериментальные результаты. Заключение и полученные результаты представлены в разделе 5.

**2. Связанные работы.** В ранних исследованиях для разработки систем распознавания жестов широко использовались одиночные камеры, так в работе [10] Т. Starner и А. Pentland представили систему, которая использует одну цветную камеру для отслеживания рук в реальном времени и интерпретирует американский язык жестов (ASL) с использованием скрытых Марковских моделей (HMM). Точность распознавания слов составляет 92%. Однако, недостатками использования одиночной камеры являются ограничение угла обзора и чувствительность к условиям освещенности, которые влияют на надежность системы [11].

Решение проблем, возникающих при распознавании 2D-изображений, привело к разработке систем, использующих в качестве сенсоров захвата стереокамеры. М. Elmezaïn и соавторы [12] предложили автоматическую систему, которая распознает, как отдельные жесты рук, так и непрерывные жесты для арабских чисел от 0 до 9 в режиме реального времени на основе скрытой Марковской модели. Входные изображения захватываются системой стереокамер Bumblebee с фокусным расстоянием 6 мм в течение примерно 2–5 секунд при 15 кадрах в секунду с разрешением изображения  $240 \times 320$  пикселей. Точность распознавания жестов составляет 98,94%. Однако, несмотря на то, что использование стереокамер позволило частично решить вопросы влияния освещения, захватывать и создавать трехмерные изображения, они имеют ряд сложностей с калибровкой и вычислением карт глубины [13].

Разработка недорогих сенсоров глубины дала новые возможности исследователям в области распознавания жестов [14]. Е. Kollorz и соавторы [15] предложили систему для распознавания 12 различных статических жестов рук на основе TOF-камеры Photonic-Mixer-Device (PMD) с разрешением  $160 \times 120$  пикселей и частотой 15 кадров в секунду. В исследовании показано, что x- и y-проекции изображения и характеристики глубины захваченного изображения достаточны для использования метода ближайших соседей. Точность распознавания системы составляет 94,61%. А. Kurakin и соавторы [16] использовали одиночную камеру Microsoft Kinect для распознавания 12 динамических жестов американского жестового языка (ASL). Для детектирования был использован метод Оцу, классификация проводилась на основе графа действий (Action Graph). Точность распознавания достигает 87,7%. Авторы работы [17] предложили систему распознавания 24 различных жестов рук, основанную на данных, полученных фронтальной камерой Intel

---

RealSense SR300. Система распознавания сопоставляет изображения глубины с цветными изображениями и распознает жесты рук с помощью сверточной нейронной сети. Точность распознавания составляет 99,4%.

M. Van den Bergh и L. Van Gool [18] провели исследования, в которых для распознавания жестов рук в реальном времени были использованы совместно камеры ToF и RGB. Классификация основана на 2D Хаарлете. Результаты показали, что такой подход, основанный на использовании данных о цвете и глубине, значительно улучшает показатели обнаружения руки - позволяет ладони перекрывать лицо и допустимо присутствие в кадре на заднем плане рук других людей.

Резюмируя краткий обзор разработанных систем, можно сделать предположение, что применение современных сенсоров глубины совместно с технологиями нейронных сетей может быть эффективным подходом для разработки систем распознавания жестов.

**3. Описание Python - приложения для распознавания жестов рук.** В данном разделе представлено описание алгоритма работы приложения, разработанного на языке Python 3.6, для распознавания жестов рук на основе предобработки видеоданных и классификации выделенных признаков при помощи нейронной сети. Для программной реализации приложения были использованы библиотеки RealSense от компании Intel, OpenCV и DL - фреймворки с открытым исходным кодом Keras и TensorFlow.

Приложение разработано с учетом использования в качестве сенсоров захвата как камеры глубины, так и камеры RGB. Отличительной особенностью работы камеры глубины от RGB камеры, является то, что в любой момент времени с камеры глубины можно получить не только цветное изображение (RGB), но и карту глубины, которая представляет собой двумерный массив чисел, каждое число которого показывает расстояние от камеры до объекта, который был спроецирован на плоскость камеры.

В приложении были использованы как методы из библиотек RealSense и OpenCV, так и методы, реализованные авторами самостоятельно. Блок-схема работы алгоритма приложения представлена на рисунке 1.

Рассмотрим более подробно работу приложения. При выборе видеисточника, пользователю доступны: веб-камера, камера компании Intel и чтение из видеофайла. Для операционных систем Windows 8 и ниже, камеру компании Intel выбрать нельзя.

Подключение к камере глубины происходит с помощью библиотеки RealSense, в которой имеются стандартные функции для инициализации камеры, установки параметров ее работы, функции и методы чтения кадров из видео потока, вычисления расстояния от руки до глубинной камеры, методы сохранения RGB-изображений и карт глубины.

Подключение к камере RGB происходит через библиотеку OpenCV, в которой также имеются необходимые функции для инициализации камеры и методы чтения кадра.

После подключения камеры Intel вычисляется расстояние до человека, определяется площадь захвата руки и вычисляется яркость первого кадра, на основании полученных результатов происходит выбор RGB кадра или карты глубины для дальнейшей работы. На следующем этапе происходит поиск области интереса (Region of Interest, ROI) на изображении для обнаружения руки и удаление фона. На основе глубинного кадра, убираются пиксели не подходящие площади захвата и путем подсчета количества не черных пикселей определяется, находится рука в кадре или нет. Если рука в кадре обнаружена, то для улучшения эффективности выполнения различных операций на последующих этапах работы системы распознавания жестов, таких как сегментация и извлечение признаков выполняется предварительная обработка изображения. На данном этапе помехи и внешние шумы уменьшаются применением операций усреднения и выравнивания гистограмм, проводится также цветовая нормализация в соответствии с условиями освещения и световой температурой, преобразование цветного изображения в оттенки серого, применение фильтра Гаусса для удаления шумов в двумерных изображениях в оттенках серого, построение контуров сегментированного объекта, пороговое преобразование для получения сегментированного жеста руки. Для распознавания жестов используется глубокая сверточная нейронная сеть (Deep Convolutional Neural Network, DCNN) с архитектурой VGG-16, предварительно обученная на большом наборе изображений, в

результате работы которой выдается распознанный жест и алгоритм повторяется для следующего кадра. Если же рука не обнаружена в кадре, считывается следующий за ним кадр.

При выборе камеры RGB, происходит подключение библиотеки OpenCV, после этого идет определение номера потока, захват заднего фона на первом кадре и его сохранение. Для фиксирования отсутствия наличия посторонних движущих объектов используется последующий кадр. Если сумма пикселей, не совпадающих с захваченным кадром, равна нулю, то данный кадр можно сохранить для последующего вычитания заднего фона, иначе фиксируются следующий кадр до тех пор, пока сумма пикселей не станет равна нулю.

На основе сохраненного кадра, убираются похожие пиксели в кадре, определяется, находится ли рука в кадре. Если рука обнаружена, то осуществляется предварительная обработка изображения, сегментация кадра и подача его на нейронную сеть для классификации. В итоге выдается распознанный жест, и работа алгоритма повторяется для следующего кадра.

На основе сохраненного кадра, убираются похожие пиксели в кадре, определяется, находится ли рука в кадре. Если рука обнаружена, то осуществляется предварительная обработка изображения, сегментация кадра и подача его на нейронную сеть для классификации. В итоге выдается распознанный жест, и работа алгоритма повторяется для следующего кадра.

**4. Экспериментальные исследования.** Апробация предложенной системы проводилась на задаче распознавания статических и динамических жестов рук, для этого была подготовлена база данных, которая содержит изображения с сегментированными жестами, представленными на рисунках 2 и 3.

В качестве устройства захвата жестов были использованы сенсоры веб-камера Logitech HD Pro Webcam C920 и камера глубины Intel RealSense D435. Технические характеристики сенсоров представлены в таблице 1.

**Таблица 1 – Технические характеристики сенсоров Logitech HD Pro Webcam C920 и Intel RealSense D435**

Модель	Logitech HD Pro Webcam C920	Intel RealSense D435
Число мегапикселей матрицы	3 Мп	3 Мп
Разрешение (видео)	1920×1080	1920×1080
Максимальная частота кадров	30 кадр./сек	90 кадр./сек
Угол обзора	77°	77°
Фокусировка	автоматическая	автоматическая
Интерфейс	USB 2.0	USB 3.0
3D-технология	-	Активный инфракрасный стерео-проектор (IR)
Выходное разрешение 3D-данных	-	До 1280×720
Частота кадров выходного 3D-потока	-	До 90 кадров в секунду
Максимальный диапазон	-	Примерно 10 метров

Распознавание статических жестов проводилось на расстояниях от камер в 25 см, 37,5 см; 43,75 см; 50 см, 62,5 см; 75 см, 100 см. В таблицах 2 и 3 представлены результаты распознавания статических жестов рук с использованием RGB- и RGBD-камер.

Сравнение полученных результатов с двух камер на расстояниях 50 см и 100 см, приведены на рисунках 4 и 5 соответственно.

Распознавание динамических жестов проводилось на расстояниях от камер в 25 см; 50 см, 75 см, 100 см. В таблицах 4 и 5 представлены результаты распознавания динамических жестов рук с использованием камер Logitech HD Pro Webcam C920 и Intel RealSense D435.

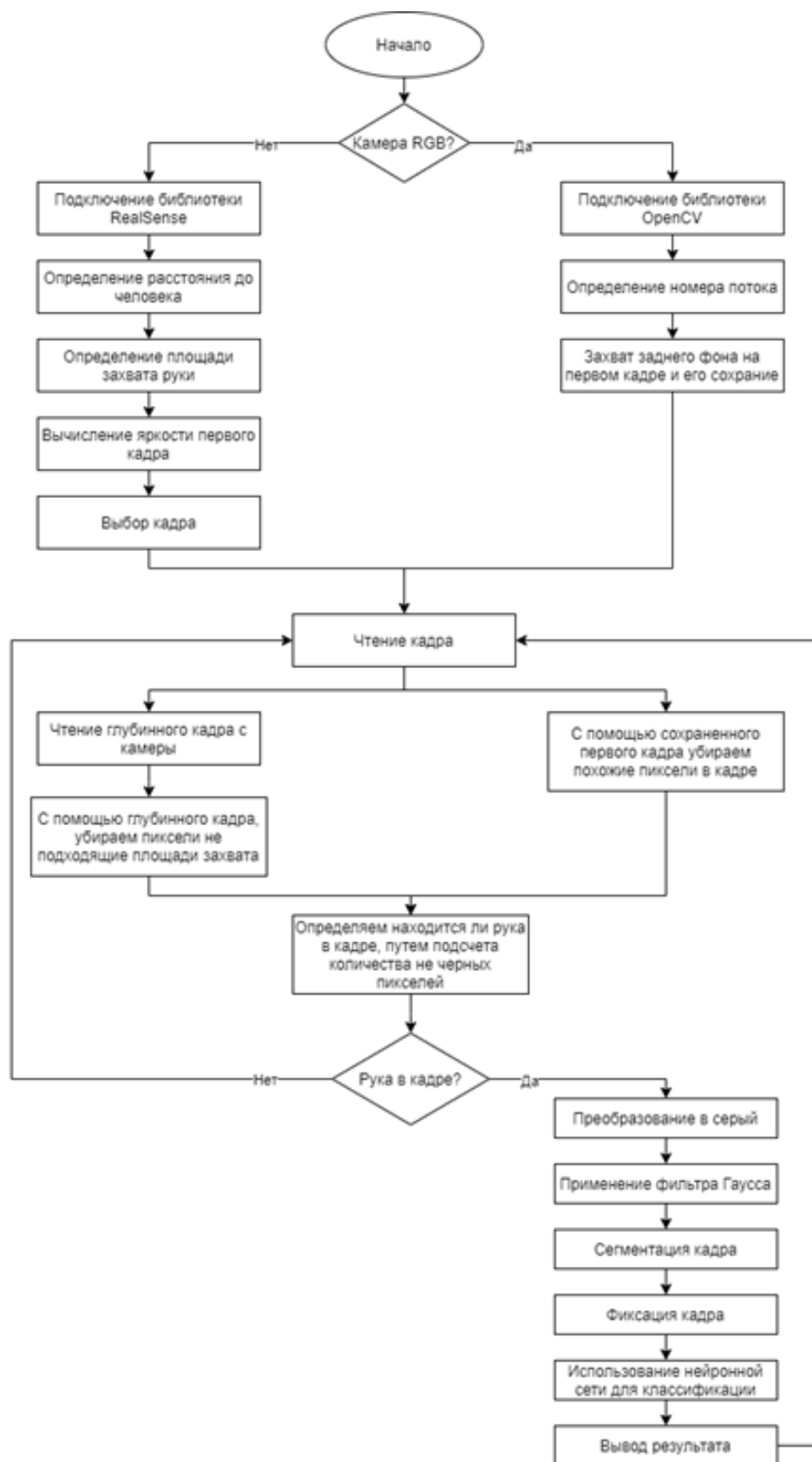


Рисунок 1 – Блок-схема алгоритма работы приложения.

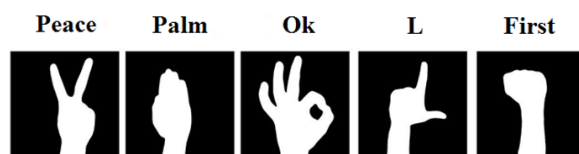


Рисунок 2 – Образцы статических жестов.

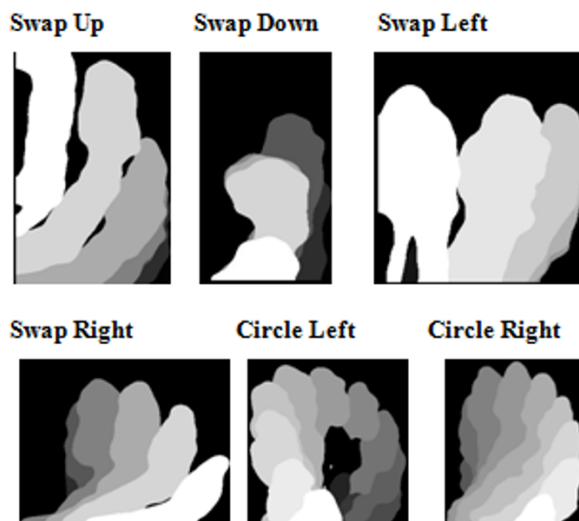


Рисунок 3 – Образцы динамических жестов.

Таблица 2 – Точность распознавания статических жестов рук на основе веб-камеры Logitech HD Pro Webcam C920

Distance, cm	Fist	L	Okay	Palm	Peace
25	98,899	99,417	99,994	77,766	10,026
37,5	98,284	99,992	99,999	99,461	22,391
43,75	99,99	99,982	100, 00	97,548	63,666
50	98,189	92,826	99,995	90,732	77,887
62,5	99,935	84,518	99,991	41,7559	61,152
75	99,972	39,8749	99,99	0,8199	91,743
100	74,114	87,447	81,246	0,022	0,202

Таблица 3 – Точность распознавания статических жестов рук на основе камеры Intel RealSense D435

Distance, cm	Fist	L	Okay	Palm	Peace
25	93,179	89,7115	99,941	49,215	10,584
37,5	97,241	97,662	86,938	99,987	82,866
43,75	99,863	99,99	99,993	99,426	76,338
50	99,685	99,991	99,993	99,981	96,025
62,5	93,025	99,975	99,99	98,253	85,407
75	11,589	99,688	89,909	3,905	4,097
100	99,547	99,281	95,091	3,06	82,973

Сравнение полученных результатов с двух камер на расстояниях 25 см и 75 см, приведены на рисунках 6 и 7 соответственно.

Распознавание статических жестов также было проведено при разных условиях освещенности, результаты распознавания приведены в таблице 6.

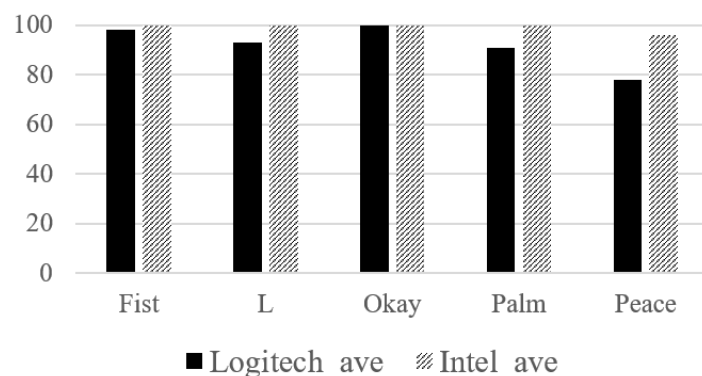


Рисунок 4 – Сравнение точности распознавания статических жестов с разных камер на расстоянии 50 см.

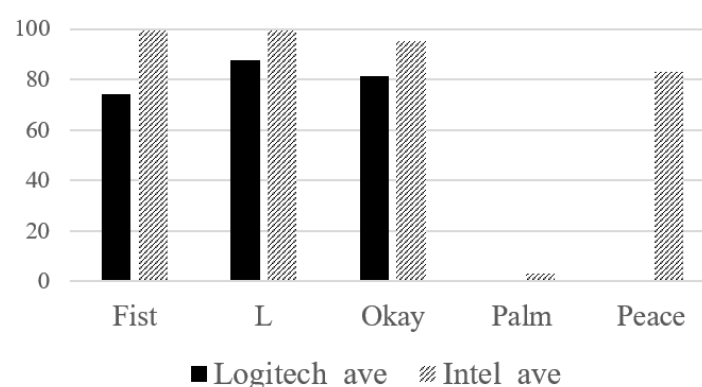


Рисунок 5 – Сравнение точности распознавания статических жестов с разных камер на расстоянии 100 см.

Таблица 4 – Точность распознавания динамических жестов рук на основе веб-камеры Logitech HD Pro Webcam C920

Distance, cm	Swap Up	Swap Down	Swap Left	Swap Right	Circle Left	Circle Right
25	0,024	99,211	98,983	90,416	99,789	50,502
50	76,2534	90,130	97,140	91,248	90,129	76,151
75	0,376	59,866	95,359	63,351	79,512	0,304
100	9,928	84,600	74,309	2,231	43,453	0,024

Таблица 5 – Точность распознавания динамических жестов рук на основе камеры Intel RealSense D435

Distance, cm	Swap Up	Swap Down	Swap Left	Swap Right	Circle Left	Circle Right
25	72,135	98,468	89,439	99,999	99,373	99,441
50	73,603	99,999	90,235	100,0	99,903	100,0
75	0,035	100,0	99,999	100,0	99,99	69,946
100	0,001	100,0	93,997	99,976	90,371	6,393

В таблицах 2-5 представлены средние вероятности классификаций статических и динамических жестов на различных расстояниях, полученных с использованием представленной системы. В качестве сенсоров захвата жестов были использованы соответственно камер Logitech HD Pro Webcam C920 и Intel RealSense D435.

Учитывая, что результирующая точность классификатора рассчитывается как среднее арифметическое его точности по всем классам, производительность системы по точности

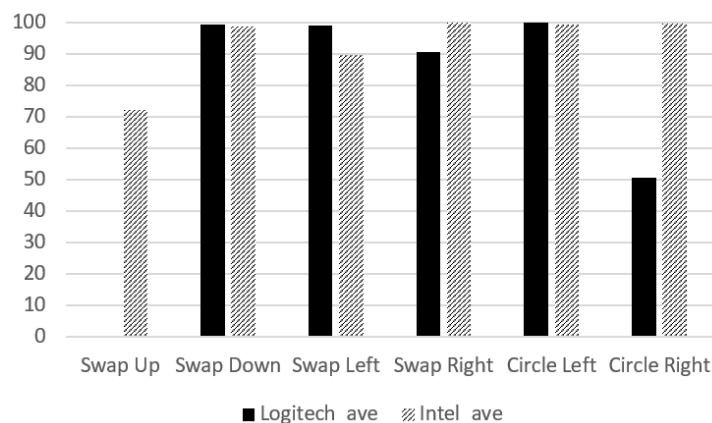


Рисунок 6 – Сравнение точности распознавания динамических жестов с разных камер на расстоянии 25 см.

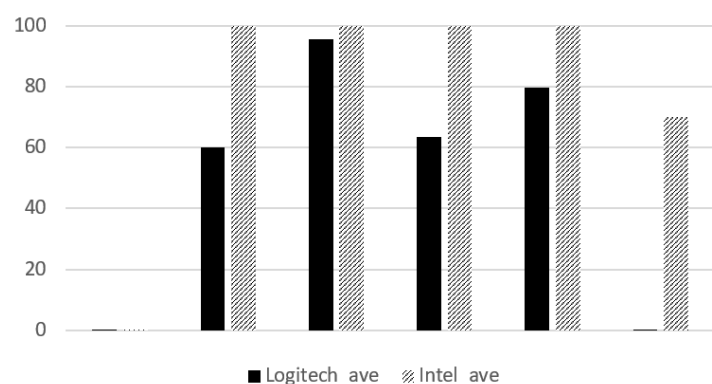


Рисунок 7 – Сравнение точности распознавания динамических жестов с разных камер на расстоянии 75 см.

Таблица 6 – Производительность системы при различных условиях освещенности для распознавания статических жестов

Условия эксперимента	Fist	L	Okay	Palm	Peace
Дневное освещение	98,015	94,713	99,984	96,399	96,74
Поздним вечером без дополнительного освещения, используются только пиксели глубины	98,542	99,899	99,729	97,629	97,838
Поздним вечером с дополнительным освещением (включенной лампочкой и светильником перед камерой)	95,577	98,308	97,305	86,271	85,324
Поздним вечером с дополнительным освещением (включенной лампочкой)	83,502	78,080	99,726	82,521	78,320
Поздним вечером без дополнительного освещения	82,522	78,072	99,402	78,548	77,120

распознавания статических жестов получена на уровне 76,76 % с использованием камеры Logitech HD Pro Webcam C920 и 81,26% с использованием изображений с камеры Intel RealSense D435. Для динамических жестов точность распознавания составила 61,38 % и 82,64% соответственно. Из полученных экспериментальных данных видно, что точность распознавания как статических, так и динамических жестов рук зависит от расстояния, на котором расположена рука от камеры, и падает с увеличением расстояния, что связано с небольшой площадью изображения руки. Исходя из полученных результатов (таблицы 2-5), оптимальное расстояние руки от камер Logitech HD Pro Webcam C920 и Intel RealSense D435 составляет 50 см, при котором для RGB камеры точность распознавания статических



---

жестов на уровне 91,93%, для RGBD камеры – 99,34%, для распознавания динамических жестов – 86,84% и 93,96% соответственно.

Как видно из Таблицы 6 производительность системы также зависит от условий освещенности в которых происходит распознавание жестов. Так при подаче на вход RGB-изображений лучший результат система показывает при дневном освещении - 98,57%, однако при использовании карт глубины производительность системы даже поздним вечером без дополнительного освещения составляет 98,73%.

Исходя из представленного анализа результатов распознавания жестов на различных расстояниях от камер, при различных условиях освещенности, а также изучения технических характеристик камер, можно сделать вывод, что распознавание жестов с использованием камеры Intel RealSense D435 дает лучшие результаты, однако использование сверточной нейронной сети позволяет получить достаточно высокую точность классификации даже в случае захвата видео с RGB-камеры.

**5. Выводы.** В работе предложен подход и программная реализация для распознавания статических и динамических жестов руки, представлены результаты экспериментальных исследований с использованием в качестве сенсоров захвата веб-камеры Logitech HD Pro Webcam C920 и камеры глубины Intel RealSense D435 при различных условиях по расстоянию и освещенности.

Показано, что точность распознавания жестов зависит от условий, при которых демонстрируются позы рук (освещенность, расстояние до камеры).

Выявлено, что использование сверточной нейронной сети позволяет получить достаточно высокую точность классификации даже в случае захвата видео с RGB-камеры. Полученная более высокая точность распознавания жестов рук с RGBD-камеры при разных условиях эксперимента подтверждает потенциальные ожидания производительности за счет пикселей глубины.

Будущие исследования связаны с практическим применением полученных результатов и разработанной программной системы для создания эффективных систем взаимодействия человека с компьютером.

## Список литературы

- 1 Liang R., Ouhyoung M. A real-time continuous gesture recognition system for sign language // IEEE Third International Conference on Automatic Face and Gesture Recognition Proceedings. – Nara, Japan, 1998. - P.558 - 567.
- 2 Vogler, C., Metaxas, D. A framework for recognizing the simultaneous aspects of American Sign Language // Computer Vision and Image Understanding - 2001. – Vol. 81. - Issue 3. - P.358-384.
- 3 Murakami K., Taguchi H. Gesture recognition using recurrent neural networks // CHI '91: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. - New Orleans, USA, 1999. - P.237-242.
- 4 Fels S. S., Hinton G. E. Glove-talk II -a neural-network interface which maps gestures to parallel formant speech synthesizer controls // IEEE transactions on neural networks – 1998. - Vol. 8. - No. 5. – P.205-212.
- 5 Zeng J., Sun Y., Wang F. A natural hand gesture system for intelligent human-computer interaction and medical assistance // Third Global Congress on Intelligent Systems – Wuhan, China, 2012. - P.382-385.
- 6 Malima A., Ozgur E., Cetin M. A fast algorithm for vision-based hand gesture recognition for robot control // IEEE 14th conference on Signal Processing and Communications Applications – Antalya, Turkey, 2006. - P.1-4.
- 7 Киселев В.А., Клименко А.С. и соавторы Аппаратные средства систем распознавания 3D жестов для комплекса виртуального окружения // Международная конференция СРТ2017. - Кипр, Ларнака, 2017. - С.152-160.
- 8 Satybaldina D., Kalymova G., Glazyrina N. Application Development for Hand Gestures Recognition with Using the Depth Camera // DB&IS 2020: Databases and Information Systems – Tallinn, Estonia, 2020. - P.55-67.
- 9 Сатыбалдина Д.Ж., Овечкин Г.В., Калымова К.А. Система распознавания статических жестов рук с использованием камеры глубины // Вестник РГРТУ – 2020. - № 72. - С.93-105.
- 10 Starner T., Pentland A. Real-time American Sign Language recognition from video using hidden Markov models // Proceedings of International Symposium on Computer Vision. - Coral Gables, USA, 1995. – P.265-270.
- 11 Howe N. R., Leventon M. E., Freeman W. T. Bayesian Reconstruction of 3D Human Motion from Single-Camera Vide // NIPS'99: Proceedings of the 12th International Conference on Neural Information Processing Systems. - 1999, P.820–826.

- 12 . Elmezain M., Al-Hamadi A., Appenrodt J., Michaelis B. A hidden Markov model-based continuous gesture recognition system for hand motion trajectory // 19th International Conference on Pattern Recognition. - Tampa, USA, 2008. - P.1– 4.
- 13 Wachs J. P., K?lsch M., Stern H., Edan Y. Vision-based hand-gesture applications // Communications of the ACM – 2011. – Vol 54. – Issue 2. – P.60 – 71.
- 14 Suarez J., Murphy R. R. Hand gesture recognition with depth images // The 21st IEEE International Symposium on Robot and Human Interactive Communication. – Paris, Franc, 2012, P.411–417.
- 15 Kollorz E., Penne J., Hornegger J., Barke A. Gesture recognition with a time-of-flight camera // International Journal of Intelligent Systems Technologies and Applications - 2008. – Vol. 5(3). – P.334-343.
- 16 Kurakin A., Zhang Z., Liu Z. A real time system for dynamic hand gesture recognition with a depth sensor // Signal Processing Conference. - Bucharest, Romania, 2012. - P.1975–1979.
- 17 Liao B., Li J., Ju Z., Ouyang G. Hand gesture recognition with generalized hough transform and DC-CNN using realsense // Eighth International Conference on Information Science and Technology. – Cordoba, Spain, 2018. – P.84-90.
- 18 Van den Bergh M., Van Gool L. Combining RGB and ToF cameras for real-time 3D hand gesture interaction // IEEE Workshop on Applications of Computer Vision. – Kona, USA, 2011. - P.66 -72.

Д.Ж. Сатыбалдина<sup>1</sup>, Н.С. Глазырина<sup>1</sup>, В.С. Степанов<sup>2</sup>, К.А. Калымова<sup>3</sup>

<sup>1</sup> Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Нұр-Сұлтан, Қазақстан

<sup>2</sup> «Kazdream Technologies» ЖШС, Нұр-Сұлтан, Қазақстан

<sup>3</sup> Қазақ ұлттық қыздар педагогикалық университеті, Алматы, Қазақстан

#### RGB және RGBD камераларының бейне ағынынан қол қимылдарды тануға арналған Python қосымшасын әзірлеу

**Аннотация:** Соңғы кезде деректерді түсіретін заманауи құрылғылардың (датчиктердің) дамуына және тану алгоритмдерінің жаңадан әзірлеуіне байланысты ым-ишараны тану жүйелері айтарлықтай өзгерді. Мақалада RGB және RGBD камераларынан, яғни Logitech HD Pro Webcam C920 веб-камерасынан және Intel RealSense D435 тереңдікті сезетін камерасынан алынған бейне ағынынан статикалық және динамикалық қол қимылдарын тануға арналған зерттеу нәтижелері келтірілген. Программалық іске асырылуы Python 3.6 құралдарының көмегімен орындалды. Python-ның ашық кітапханалары кескіндерді өңдеу мен сегменттеу алгоритмдерін сенімді түрде жүзеге асырады. Қимылдарды бөліп алу және жіктеу ішжүйесі TensorFlow және Keras терең оқыту фреймворкелер көмегі арқылы жүзеге асырылған, VGG-16 нейрондық желісінің архитектурасына негізделген. Камералардың техникалық сипаттамалары келтірілген. Қосымшаның жұмыс тәртібі сипатталған. Әртүрлі эксперименттік жағдайлардағы (қашықтықты және жарықтандыру өзгерту арқылы) деректерді түсіретін құрылғыларды салыстыруға бағытталған зерттеулер ұсынылған. Эксперимент нәтижелері көрсеткендей, Intel RealSense D435 тереңдік камерасы әртүрлі эксперименттік жағдайларда қимылдарды дәлірек тануға мүмкіндік береді.

**Түйін сөздер:** тереңдік камерасы, қимылдарды тану, орамдық нейрон желі, RealSense, OpenCV, Python, VGG-16.

D.Zh. Satybalдина<sup>1</sup>, N.S. Glazyrina<sup>1</sup>, V.S. Stepanov<sup>2</sup>, K.A. Kalymova<sup>3</sup>

<sup>1</sup> L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan

<sup>2</sup> «Kazdream Technologies» LLP, Nur-Sultan, Kazakhstan

<sup>3</sup> Kazakh national women's teacher training university, Almaty, Kazakhstan

#### Development of a Python application for recognizing gestures from a video stream of RGB and RGBD cameras

**Abstract:** Gesture recognition systems have changed a lot recently, due to the development of modern data capture devices (sensors) and the development of new recognition algorithms. The article presents the results of a study for recognizing static and dynamic hand gestures from a video stream from RGB and RGBD cameras, namely from the Logitech HD Pro Webcam C920 webcam and from the Intel RealSense D435 depth camera. Software implementation is done using Python 3.6 tools. Open source Python libraries provide robust implementations of image processing and segmentation algorithms. The feature extraction and gesture classification subsystem is based on the VGG-16 neural network architecture implemented using the TensorFlow and Keras deep learning frameworks. The technical characteristics of the cameras are given. The algorithm of the application is described. The research results aimed at comparing data capture devices under various experimental conditions (distance and illumination) are presented. Experimental results show that using the Intel RealSense D435 depth camera provides more accurate gesture recognition under various experimental conditions.

**Keywords:** depth camera, gesture recognition, convolutional neural network, RealSense, OpenCV, Python, VGG-16.

## References

- 1 Liang R., Ouhyoung M. A real-time continuous gesture recognition system for sign language, IEEE Third International Conference on Automatic Face and Gesture Recognition Proceedings.Nara, Japan, 1998. P. 558-567.
- 2 Vogler, C., Metaxas, D. A framework for recognizing the simultaneous aspects of American Sign Language, Computer Vision and Image Understanding, 81(3), 358-384(2001).

- 3 Murakami K., Taguchi H. Gesture recognition using recurrent neural networks, CHI '91: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New Orleans, USA, 1999. P.237-242.
- 4 Fels S. S., Hinton G. E. Glove-talk II -a neural-network interface which maps gestures to parallel formant speech synthesizer controls, IEEE transactions on neural networks, 8(5), 205-212(1998).
- 5 Zeng J., Sun Y., Wang F. A natural hand gesture system for intelligent human-computer interaction and medical assistance, Third Global Congress on Intelligent Systems. Wuhan, China, 2012. P.382-385.
- 6 Malima A., Ozgur E., Cetin M. A fast algorithm for vision-based hand gesture recognition for robot control, IEEE 14th conference on Signal Processing and Communications Applications – Antalya, Turkey, 2006. P.1-4.
- 7 Kiselev V. A., Klimenko A. S. and co-authors Apparatsnye sredstva sistem raspoznavaniya 3D zhestov dlya kompleksa virtual'nogo okruzheniya [3D Gesture recognition hardware tools for the Virtual Environment system], Mezhdunarodnaya konferenciya CPT2017[International conference CPT2017]. Larnaca, Cyprus, 2017. P.152-160. [in Russian].
- 8 Satybaldina D., Kalymova G., Glazyrina N. Application Development for Hand Gestures Recognition with Using the Depth Camera, DB&IS 2020: Databases and Information Systems. Tallinn, Estonia, 2020. P.55-67.
- 9 Satybaldina D.Zh., Ovechkin G.V., Kalymova G.A. Sistema raspoznavaniya staticheskikh zhestov ruk s ispol'zovaniem kamery glubiny [Static hand gestures recognition system with using depth camera], Vestnik RGRU[Vestnik of RSREU], N. 72, 93-105(2020). [in Russian].
- 10 Starner T., Pentland A. Real-time American Sign Language recognition from video using hidden Markov models, Proceedings of International Symposium on Computer Vision. Coral Gables, USA, 1995. P.265-270.
- 11 Howe N. R., Leventon M. E., Freeman W. T. Bayesian Reconstruction of 3D Human Motion from Single-Camera Vide, NIPS'99: Proceedings of the 12th International Conference on Neural Information Processing Systems. 1999, P.820–826.
- 12 Elmezain M., Al-Hamadi A., Appenrodt J., Michaelis B. A hidden Markov model-based continuous gesture recognition system for hand motion trajectory // 19th International Conference on Pattern Recognition. Tampa, USA, 2008. P.1– 4.
- 13 Wachs J. P., K?lsch M., Stern H., Edan Y. Vision-based hand-gesture applications, Communications of the ACM. 2011. Vol 54. Issue 2. P.60-71.
- 14 Suarez J., Murphy R. R. Hand gesture recognition with depth images, The 21st IEEE International Symposium on Robot and Human Interactive Communication. Paris, Franc, 2012, P.411–417.
- 15 Kollorz E., Penne J., Hornegger J., Barke A. Gesture recognition with a time-of-flight camera, International Journal of Intelligent Systems Technologies and Applications, 5(3), 334-343(2008).
- 16 Kurakin A., Zhang Z., Liu Z. A real time system for dynamic hand gesture recognition with a depth sensor, Signal Processing Conference. Bucharest, Romania, 2012. P.1975–1979.
- 17 Liao B., Li J., Ju Z., Ouyang G. Hand gesture recognition with generalized hough transform and DC-CNN using realsense, Eighth International Conference on Information Science and Technology. Cordoba, Spain, 2018. P.84-90.
- 18 Van den Bergh M., Van Gool L. Combining RGB and ToF cameras for real-time 3D hand gesture interaction, IEEE Workshop on Applications of Computer Vision. Kona, USA, 2011. P.66 -72.

**Сведения об авторах:**

*Сатыбалдина Д.Ж.* - **автор для корреспонденции**, кандидат физико-математических наук, ассоциированный профессор, заведующий кафедрой «Информационная безопасность» Евразийского национального университета им. Л.Н. Гумилева, Пушкина 11, Нур-Султан, Казахстан.

*Глазырина Н.С.* - PhD, и.о. доцента кафедры «Компьютерная и программная инженерия» Евразийского национального университета им. Л.Н. Гумилева, Пушкина 11, Нур-Султан, Казахстан.

*Степанов В.С.* – магистр, ТОО «Kazdream Technologies», Туркестан 10, Нур-Султан, Казахстан.

*Калымова К.А.* – магистр, старший преподаватель кафедры «Информатики и прикладной математики» Казахского национального женского педагогического университета, Айтеке би 99, Алматы, Казахстан.

*Satybaldina D.Zh.* - **corresponding author**, candidate of physical and mathematical sciences, Associate Professor, Head of the Department of Information Security, at L.N. Gumilyov Eurasian National University, Pushkin 11, Nur-Sultan, Kazakhstan.

*Glazyrina N.S.* - PhD, Associate Professor of the Department of Computer and Software Engineering at L.N. Gumilyov Eurasian National University, Pushkin 11, Nur-Sultan, Kazakhstan.

*Stepanov V.S.* - Master, «Kazdream Technologies» LLC, Turkestan 10, Nur-Sultan, Kazakhstan.

*Kalymova G.A.* - Master, Senior Lecturer of the Department of Informatics and Applied Mathematics, Kazakh National Women's Teacher Training University, Aiteke bi 99, Almaty, Kazakhstan.