УДК 004.8
DOI: 10.25559/SITITO.16.202004.883-892

Оригинальная статья

# Искусственный интеллект в задачах распознавания динамических жестов

**Е. Р. Муратов, М. Б. Никифоров*, А. М. Скачков**
ФГБОУ «Рязанский государственный радиотехнический университет имени В.Ф. Уткина», г. Рязань, Российская Федерация
390005, Российская Федерация, г. Рязань, ул. Гагарина, д. 59/1
* nikiforov.m.b@mail.ru

## Аннотация

Разработка эффективных методов распознавания жестов руки человека является актуальной задачей как с научной, так и с прикладной точки зрения. Методы распознавания жестов лежат в основе бесконтактных интерфейсов управления техническими системами. Наиболее значимыми областями применения систем распознавания жестов являются автоматический сурдоперевод и бесконтактное управление техническими системами. Почти все существующие реализации методов распознавания уверенно работают, когда рука находится на однородном фоне. Но в реальности такой случай использования подобных реализаций маловероятен. Вероятность правильного распознавания резко снижается на сложном фоне, когда рука на изображении пересекает лицо или элементы тела с открытой кожей. Дополнительные ограничения на алгоритмы накладывают требования к их аппаратной реализации. Это должны быть компактные устройства на кристалле (SoC) с малым потреблением, габаритами и ценой и, следовательно, с малой вычислительной мощностью. Учитывая сказанное, следует признать актуальным решения задачи по повышению эффективности алгоритмов и методов распознавания жестов.
Рассматриваются два подхода к решению задачи распознавания динамических жестов руки - аналитический и нейросетевой. Показано, что использование приемов искусственного интеллекта может повысить достоверность распознавания жестов в сложных условиях видеонаблюдения. Однако, применение нейросетевых алгоритмов не может показать высокую производительность на одноплатных компьютерах, если они не имеют NPU или производительного GPUмодуля.

**Ключевые слова:** искусственный интеллект, аналитический алгоритм, статические и динамические жесты, бесконтактное управление техническими системами, сложный фон, распознавание.

*Авторы заявляют об отсутствии конфликта интересов.*

**Для цитирования:** Муратов, Е. Р. Искусственный интеллект в задачах распознавания динамических жестов / Е. Р. Муратов, М. Б. Никифоров, А. М. Скачков. – DOI 10.25559/SITITO.16.202004.883-892 // Современные информационные технологии и ИТ-образование. – 2020. – Т. 16, № 4. – С. 883-892.

Original article

# Artificial Intelligence in Recognition of Dynamic Gestures

**Ye. R. Muratov, M. B. Nikiforov*, A. M. Skachkov**
Ryazan State Radio Engineering University named after V.F. Utkin, Ryazan, Russian Federation
59/1 Gagarin St., Ryazan 390005, Russian Federation
* nikiforov.m.b@mail.ru

## Abstract
The development of effective methods for recognizing human hand gestures is an urgent task, both from a scientific and an applied point of view. Gesture recognition methods are the basis of contactless interfaces for managing technical systems. The most significant areas of application of gesture recognition systems are automatic sign language translation and contactless control of technical systems. Almost all existing implementations of recognition methods work confidently when the hand is on a uniform background. But in reality, such a case of using such implementations is unlikely. The likelihood of correct recognition drops sharply against a complex background when the hand in the image crosses the face or body parts with exposed skin. Additional restrictions on algorithms impose requirements for their hardware implementation. These should be compact devices on a chip (SoC) with low consumption, size and price, and, therefore, with low computing power. Taking into account the above, it should be recognized that it is relevant to solve the problem of increasing the efficiency of algorithms and methods of gesture recognition.
Two approaches to solving the problem of recognizing dynamic hand gestures are considered - analytical and neural network. It is shown that the use of artificial intelligence techniques can increase the reliability of gesture recognition in complex video surveillance conditions. However, the use of neural network algorithms cannot show high performance on single-board computers if they do not have an NPU or a powerful GPU module.

**Keywords:** artificial intelligence, analytical algorithm, static and dynamic gestures, contactless control of technical systems, complex background, recognition.

*The authors declare no conflict of interest.*

**For citation:** Muratov Ye.R., Nikiforov M.B., Skachkov A.M. Artificial Intelligence in Recognition of Dynamic Gestures. *Sovremennye informacionnye tehnologii i IT-obrazovanie* = Modern Information Technologies and IT-Education. 2020; 16(4):883-892. DOI: https://doi.org/10.25559/SITITO.16.202004.883-892

## I. Актуальность проблемы

Разработка эффективных методов распознавания жестов руки человека является актуальной задачей, как с научной, так и с прикладной точки зрения. Методы распознавания жестов лежат в основе бесконтактных интерфейсов управления техническими системами[1,2] [1,2]. Существуют внедренные, простые реализации (интерфейс управления мультимедиа системой в некоторых моделях автомобилей) на базе датчиков глубины, но работающие на расстоянии нескольких сантиметров и распознающие определенные движения руки относительно датчика [3]. Наиболее значимыми областями применения систем распознавания жестов являются автоматический сурдоперевод и бесконтактное управление техническими системами[3] [4,5,6,7,8,9,10,11]. В тоже время большинство известных алгоритмов и методов[4] [12,13,14,15,16,17,18,19,20] нельзя признать полностью удовлетворяющими потребителей. Почти все существующие реализации методов распознавания уверенно работают, когда рука находится на однородном фоне. Это дает высокую степень уверенности нахождения руки на изображении [5,21,22,23]. Но в реальности такой случай использования подобных реализаций маловероятен. Вероятность правильного распознавания резко снижается на сложном фоне, когда рука на изображении пересекает лицо или элементы тела с открытой кожей, что является наиболее распространенной ситуацией, когда пользователь смотрит в сторону фиксирующего жест сенсора. Почти все алгоритмы обучены или настроены на детектирование ладони, пальцев и других элементов руки, изображенных в кадре перпендикулярно плоскости сенсора [16,24,25,26]. Если сенсор видит руку под другим углом, это не позволяет алгоритмам детектировать ладонь (как и собственно жест) по последовательности изображений. Также на достоверность распознавания сильно влияют перчатки, перстни и др., как правило, не участвующие в настройке алгоритмов аксессуары. Дополнительные ограничения на алгоритмы накладывают требования к их аппаратной реализации. Это должны быть компактные устройства на кристалле (SoC) с малым потреблением, габаритами и ценой, а, следовательно, и с малой вычислительной мощностью. Учитывая сказанное, следует признать актуальным решения задачи по повышению эффективности алгоритмов и методов распознавания жестов.

## II. Видеокомпьютерные системы распознавания жестов

Жесты можно условно разделить на статические и динамические. Методы их распознавания также можно разделить на два класса – аналитические и нейросетевые.

Статические жесты [13,24,27] следует определить следующим образом: в течение некоторого фиксированного временного интервала рука неподвижна, и в это время происходит фиксация и считывание изображения руки. В этом случае задачей системы является анализ конфигурации ладони и сравнение её с заданным шаблоном.

Динамические жесты подразделяются на простые и сложные [26,28,29,30]. Простые представляют собой 2, 3, 4 фиксированных положения руки (чаще 2). Задачей системы распознавания в этом случае является фиксация факта перехода из одного положения в другое и идентификация каждого из них. Наиболее трудны для распознавания сложные динамические жесты. В них необходимо не только идентифицировать фиксированные положения, но и траектории движения рук. Существующие реализации уверенного распознавания жестов рук требуют использования ToF сенсоров или сенсоров глубины на базе структурированного подсвета [4,25,28,31,32]. Существуют реализации методов, основанных на базе анализа цвета и формы с помощью разных моделей ладони и пальцев руки[23,24,33]. В последнее время наблюдается тенденция в решении подобных задач нейросетевыми методами. Существует множество нейросетевых методов. Самые известные: MediaPipe [14] от Google, детектирование руки с использованием сетей CNN на Tensorflow [20,25,34,35,36], а также множество реализаций типа игры камень-ножницы-бумага (rock-paper-scissor) [35]. Следует отметить ограниченность применения существующих реализаций. Например, реализация от Google не работает, если рука находится в перчатке. Также неустойчивая работа наблюдается, когда рука в кадре перекрывает лицо. Жест сжатия руки в кулак для нейросетевого детектора является одним из самых сложных. Часто, когда жест выполняется на некотором удалении от камеры, системы технического зрения классифицирует лицо как кулак. В реальности устойчивая работа наблюдается только на почти однородном фоне. Анализ существующих реализаций показал, что сложность используемых в прототипах архитектур нейросетей не позволяет реализовать ее на существующих дешевых одноплатных системах с ARM архитектурой. Отсутствие достаточно производительной GPU или аппаратного нейросетевого вычислительного блока (NPU) являются препятствием для реализации существующих нейросетевых алгоритмов в реальном времени. Существуют различные методы на базе нахождения и анализа контура ладони, но все эти методы достаточно ресурсоемкие. В отличие от нейросетевых реализаций менее ресурсозатратным является эвристический анализ. Возможно распознавание объекта на основе статистики изменения ключевых характеристик объекта. Такой подход будет давать результат при условии, что удастся соотнести ключевые элементы изображения конкретным объектам в кадре и осуществить их трекинг.

---

[1] *Абакумов В. Г., Ломакина Е. Ю.* Автоматическое распознавание жестов в интеллектуальных системах [Электронный ресурс] // Искусственный интеллект. 2010. № 3. С. 269-273. URL: http://dspace.nbuv.gov.ua/bitstream/handle/123456789/56273/31-Abakumov.pdf (дата обращения: 21.08.2020).

[2] APDS-9960. Digital Proximity, Ambient Light, RGB and Gesture Sensor [Электронный ресурс]. URL: https://cdn.sparkfun.com/assets/learn_tutorials/3/2/1/Avago-APDS-9960-datasheet.pdf (дата обращения: 21.08.2020).

[3] *Brenner H.* Training a Neural Network to Detect Gestures with OpenCV in Python [Электронный ресурс] // Towards Data Science. URL: https://towardsdatascience.com/training-a-neural-network-to-detect-gestures-with-opencv-in-python-e09b0a12bdf1 (дата обращения: 21.08.2020).

[4] *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. DOI: https://doi.org/10.21667/1995-4565-2020-72-93-105

В данной работе рассматривается система распознавания динамического жеста, состоящего из двух фиксированных положений (рис. 1):
- раскрытая ладонь,
- ладонь, сжатая в кулак.

Р и с. 1. Динамический жест
F i g. 1. Dynamic gesture

Применительно к рассматриваемому жесту авторами предлагается два варианта эффективных алгоритмов: аналитический, основанный на поиске и анализе движения ключевых точек изображения руки и нейросетевой, использующий нейронную сеть семейства SSD (SingleShotDetector).

## III. Предлагаемый аналитический алгоритм

Алгоритм детектирования жеста сводится к нахождению ключевых точек на изображении [1, 37]: для сокращения вычислительной сложности повышения качества распознавания предлагается применить детектор углов методом Shi&Tomasi [38]. Так как во время наблюдения жеста пиксели, принадлежащие руке, перемещаются по изображению, предлагается фильтровать результат детектирования углов с помощью маски движения. Отфильтрованные вершины углов с помощью приёма кластеризации объединяются в объекты – кластеры, которым присваиваются номера. Значение номера кластера удерживается при его перемещении на изображении от кадра к кадру. Таким образом, детектирование жеста руки сводится к анализу изменения ключевых характеристик кластера: высота и ширина описывающего прямоугольника, количество детектированных вершин углов, положение центра на изображении, событие распадения или объединение кластера из нескольких других. В результате анализа можно с высокой достоверностью определить события сжатия ладони в кулак и разжатие кулака. И если эти два события наступают с небольшим интервалом времени, можно судить о фиксации жеста «Сжатие в кулак/разжатие». В реальности кадры, поступающие от видеокамеры, должны пройти предварительную обработку. Цель предобработки — получить вектор вершин углов детектора углов методом Shi&Tomasi [38], а также уменьшить влияние шума на результат детектирования углов. Каждый кадр от видеокамеры представляет собой изображение в форматеRGB. С целью сокращения объема обрабатываемой информации для каждого кадра видеоряда $F_n$ в формате RGB, выделяется оптическая характеристика — яркость *Y*, путем перевода RGB изображения в цветовое пространство YUV

$$\begin{bmatrix} Y' \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.14713 & -0.28886 & 0.436 \\ 0.615 & -0.51499 & -0.10001 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix},$$

где: *Y* — яркость объектов на изображении, *U* и *V* цветоразностные компоненты. Затем изображение, состоящее только из компоненты *Y,* сворачивается с ядром Гаусса размером 5x5 пикселей

$I=g(Y, \sigma)$,

где: g — функция гаусса, σ — параметр размытия обеспечивающий шумоподавление. Результатом детектирования углов на изображении *I* является *C* — вектор координат пикселей, которые детектор принял за вершины углов границ объектов на изображении *I*. Для фильтрации вершин углов принадлежащих только движущимся объектам применяется маска M вычисляемая по текущему и предыдущему предварительно обработанными кадрами

$M= \|\nabla I_{n-1}| - |\nabla I_n\| > \varepsilon$ ,

где: $|\nabla I|$ — модуль градиента, вычисленный методом Превитта, *n*- номер кадра, ε — порог яркости для построения маски (предлагается значение ε=10).

В результате фильтрации вектора координат вершин *C* получим новый вектор вершин C', для которых значение в маске *M* есть истина. На самом деле, вектор C' содержит координаты пикселей, являющихся проекцией на плоскость камеры вершин углов, границ объектов. Поэтому элементы из *C'* будем далее называть координатами пикселей.

## Кластеризация

Алгоритм кластеризации состоит из следующих шагов.

1. Вычисление *E* — области изображения для каждого элемента вектора *C'{x,y}*

$$E_i = \forall I(x', y'),$$
$$x' \in \left[C_i'\{x\} - R, C_i'\{x\} + R\right], (x', y'),$$
$$y' \in \left[C_i'\{y\} - R, C_i'\{y\} + R\right],$$

где: *i* — номер элемента вектора *C', x'* и *y'* координаты пикселя, *R* — значение для построения прямоугольной области с центром в координатах .

2. Выполнение первичной кластеризации объектов *O'* путем объединения всех пересекающихся областей *E*

$$O' = \bigcup E, \forall E_i \bigcup E_j \neq 0.$$

3. Полученные *m* объектов *O'* образуют окончательный кластер *O* путем объединения

$$O = O' \bigcup O'' / O''',$$

где:*O''*- кластеры (объекты) полученные при обработке пре-

дыдущего кадра видеоряда, O''' — кластеры полученные при обработке *i-k* кадра, *i* – номер текущего кадра, *k*- длина анализируемой последовательности кадров. В результате такого объединения кластеру присваивается номер наибольшего по количеству занимаемых пикселей на изображении кластера с предыдущего кадра и вошедшему в состав текущего кластера путем объединения. Оставшимся кластерам для текущего кадра номера присваиваются таким образом, чтобы они не повторялись для всех кластеров текущего изображения.

## Определение жеста руки

Алгоритм определяет жест сжатия/расжатия кулака как два последовательных события: сжатие ладони в кулак (начало жеста DOWN) и расжатие кулака в ладонь (конец жеста UP), разделенные небольшим временным промежутком (не более 1 секунды). Алгоритм состоит из следующих шагов.

1. Вычисление параметров кластеров для текущего кадра с номером *n*:

- $W_n(O_k)$ – ширина прямоугольника, описывающего кластер с номером *k*;
- $H_n(O_k)$ – высота прямоугольника, описывающего кластер с номером *k*;
- $C_n(O_k)$ – количество вершин углов, образовавших кластер с номером *k;*
- $S_n(O_k)$ – количество пикселей, занимаемых кластером с номером *k;*
- $P_n(O_k)=S_n(O_k)/(W_n(O_k)*H_n(O_k)$;
- $F_n(O_k)$ – прямоугольная область изображения, нижнее ребро которого является одновременно верхним ребром прямоугольника, описывающего кластер $O_k$, а высота боковых ребер составляет ~60% от высоты бокового ребра прямоугольника, описывающего кластер $O_k$;
- $D_n(O_k)$ – количество других кластеров , пересекающих область $F_n(O_k)$;
- $Y_n(O_k) = H_n(O_k)+min(H_n(O_m)$;
- $U_n(O_k)$ — количество кластеров с предыдущего кадра, которые объединились в один кластер $O_k$ на текущем кадре *n*.

2. Фиксация события DOWN, если наступают условия:

а) $(C_{n-1}(O_k)>0$ *and* $C_n(O_k)>5)| C_n(O_k)>16$;

б) $C_{n-1}(O_k)<C_n(O_k)$ — увеличивается количество детектированных на объекте вершин углов;

в) $H_n(O_k)/ W_n(O_k)<\delta_{hw}$ – , где $\delta_{hw}$ пороговое значение (выбрано как 0,8) — кластер имеет определенное отношение высоты и ширины;

г) $H_{n-1}(O_k)>H_n(O_k)$ *or* $U_n(O_k)>1$ – либо высота кластера уменьшается, либо кластер сформирован из объединения нескольких кластеров с предыдущего кадра;

д) $P_{n-1}(O_k)<P_n(O_k)$ *or* $U_n(O_k)>1$ – отношение количество пикселей объекта к площади описывающего прямоугольника увеличивается, либо кластер сформирован из объединения нескольких кластеров с предыдущего кадра;

е)$P_{n-1}(O_k)>0 | P_n(O_k)>\delta_p$ – это условие ограничивает наступление события DOWN, если кластер только что образовался на текущем кадре или имеет достаточно высокое отношение количества пикселей к площади описывающего прямоугольника;

ж) Выполнение условий а)-е) фиксируется к кадрам подряд, где к – длина анализируемой последовательности видеоряда (предлагается значение к=5).

3. Фиксация события UP, если наступают условия:

а) $(C_{n-1}(O_k)>0$ *&&*$C_n(O_k)>3)$;

б) $C_{n-1}(O_k)>C_n(O_k)$ — уменьшается количество вершин углов в кластере;

в) $H_{n-1}(O_{k-1})<Y_n(O_k)$– высота кластера увеличивается, при этом учитывается высота кластеров, которые могут представлять пальцы руки и находиться выше ладони;

г) $H_{n-1}(O_k)<H_n(O_k)$ *or* $D_n(O_k)>1$;

д) $P_n(O_{k-1})>P_n(O_k)$ *or* $D_n(O_k)>1$;

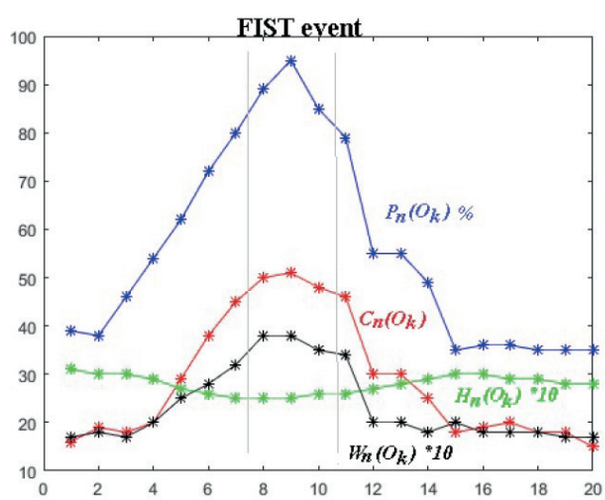
е) Выполнение условий а)-д) фиксируется к кадрам подряд, где к – длина анализируемой последовательности видеоряда (предлагается значение к=5).

4. Фиксация жеста при условии наступления события DOWN до события UP и между двумя прошло достаточно малое время, менее 1 секунды.

Условия подобраны из соображения маловероятности их выполнения, если кластер представляет какой-то другой объект.

## Результаты

На рисунке 2 представлена типовая зависимость изменения значений статистик для случая сжатия ладони в кулак.



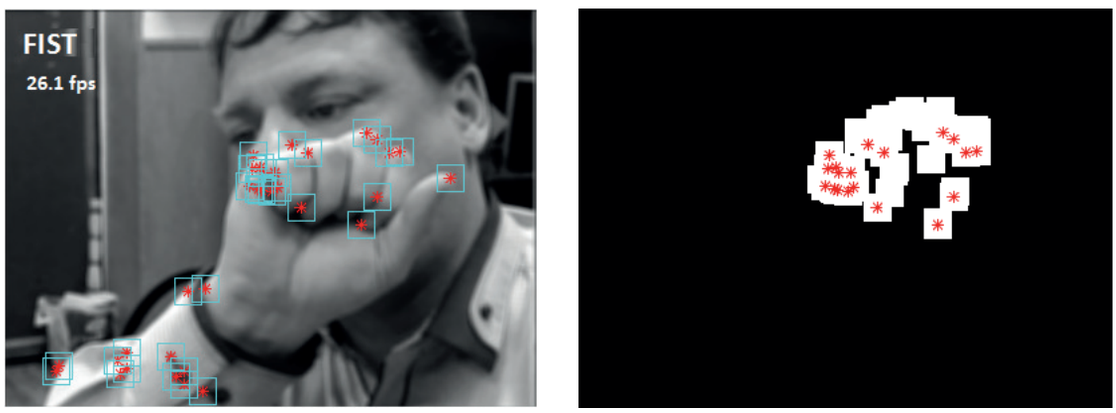

Р и с. 2. Типичное изменение значений некоторого кластера, характеризующих наступление состояния «сжатие ладони в Кулак»

F i g. 2. Typical change in the values of a certain cluster, characterizing the onset of the state of "clenching the palm into a fist"

Как видно из зависимостей, в момент сжатия ладони в кулак увеличивается количество детектированных вершин, объединенных в один кластер, наблюдается рост ширины кластера, локальная концентрация вершин порождает рост отношения количества пикселей, занимаемых окрестностями детектированных вершин к площади кластера.

На рисунке 3 представлен результат работы алгоритма кластеризации. На рисунке 3а) отмечены вершины углов, обнаруженные детектором для подвижных объектов в текущем кадре. На рисунке 3б) представлен один из образованных кластеров, полученных в результате объединения разных кластеров с пре-

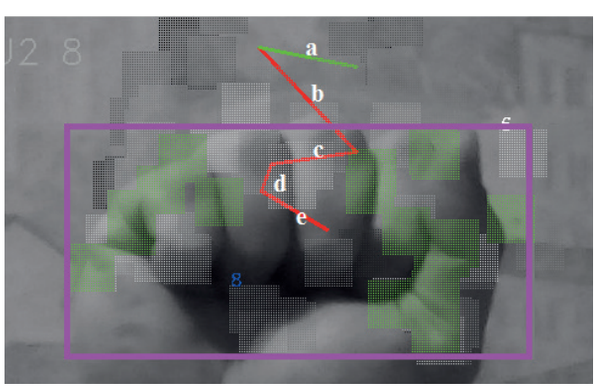
дыдущих кадров. Также рисунок 3а демонстрирует фиксирование события «Сжатие ладони» алгоритмом фиксации жеста. Алгоритм кластеризации позволяет осуществить простейший трекинг объекта и отследить, какие кластеры объединяются в один, какие — распадаются на несколько отдельных, а также, как кластеры перемещаются в поле зрении камеры.




а) б)

Р и с. 3 – a) Отфильтрованный результат детектора Shi&Tomasi и детектированное событие «Fist»,
b) сегмент и его элементы, принадлежащие объекту «кулак»
F i g. 3 – a) Filtered Shi & Tomasi detector result and detected "Fist" event,
b) segment and its elements belonging to the "Fist" object

В момент совершения жеста, на примере сжатия ладони в кулак наблюдается смещение центра кластера, что является следствием уменьшения его высоты (см. рисунок 4).




Р и с. 4. abcde – вектор перемещения центра кластера на последовательности из 5 кадров
F i g. 4. abcde – vector of displacement of the cluster center on a sequence of 5 frames

Анализ изменения значений измеренных характеристик кластера для последовательности соседних кадров позволяет достоверно детектировать события, сжатие ладони в кулак и расжатие. Одними из важных характеристик являются: изменение количества детектированных углов в отслеживаемом сегменте, а также признак объединения или распадение кластера на несколько более мелких. Результат тестирования реализации алгоритма детектирования жеста на CPU IntelAtom показал производительность на уровне 25 FPS.
Тестирование реализации показало возможность использования дешевых и энергоэффективных вычислительных ресурсов в том числе и SoC для аппаратной реализации интерфейса управления жестами.
Сравнение результатов тестирования предложенного аналитического алгоритма с реализацией MediaPipe представлены в таблице 1.

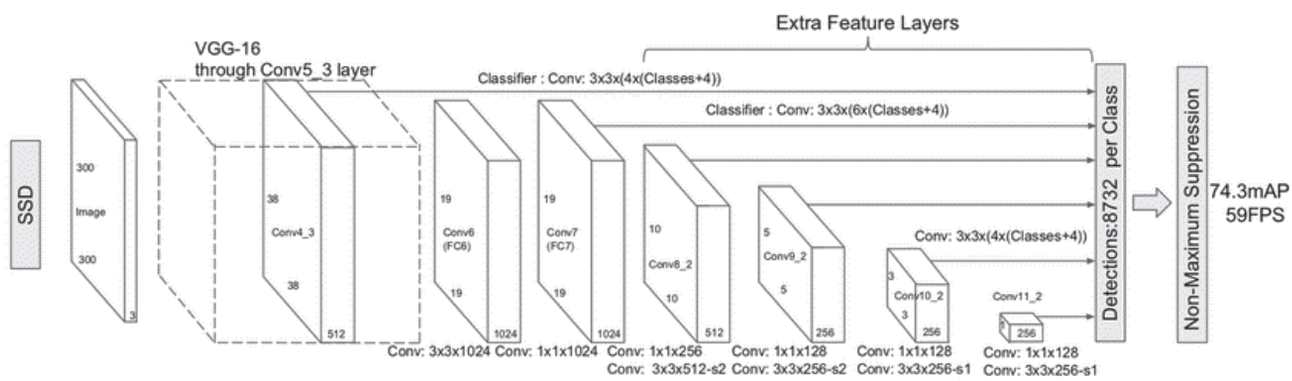
Т а б л и ц а 1. **Сравнение предложенного метода с MediaPipe**
T a b l e 1. **Comparison of the proposed method with MediaPipe**

| | **MediaPipe** | **Предложенный алгоритм** |
|---|---|---|
| Потребление памяти | 12Мб | Менее 1 Мб |
| Кол-во кадров на IntelAtom Z8500 | 1 FPS | 30 FPS |
| Процент правильного детектирования ладони (ладонь занимает 60% от высоты кадра) | 98% | 90% |
| Процент правильного детектирования ладони в перчатке (ладонь занимает 60% от высоты кадра) | 6% | 83% |
| Процент правильного детектирования ладони (ладонь занимает 40% от высоты кадра) | 93% | 83% |
| Процент правильного детектирования ладони в перчатке (ладонь занимает 40% от высоты кадра) | 3% | 81% |
| Процент правильного детектирования жеста «кулак» (занимает 20% от высоты кадра) | 45% | 24% |
| Процент ошибочного детектирования «Кулак» при размере объекта 60% от высоты кадра | 3% | 6% |
| Количество детектируемых жестов | Обучается сетью | 5-6 разных жестов |

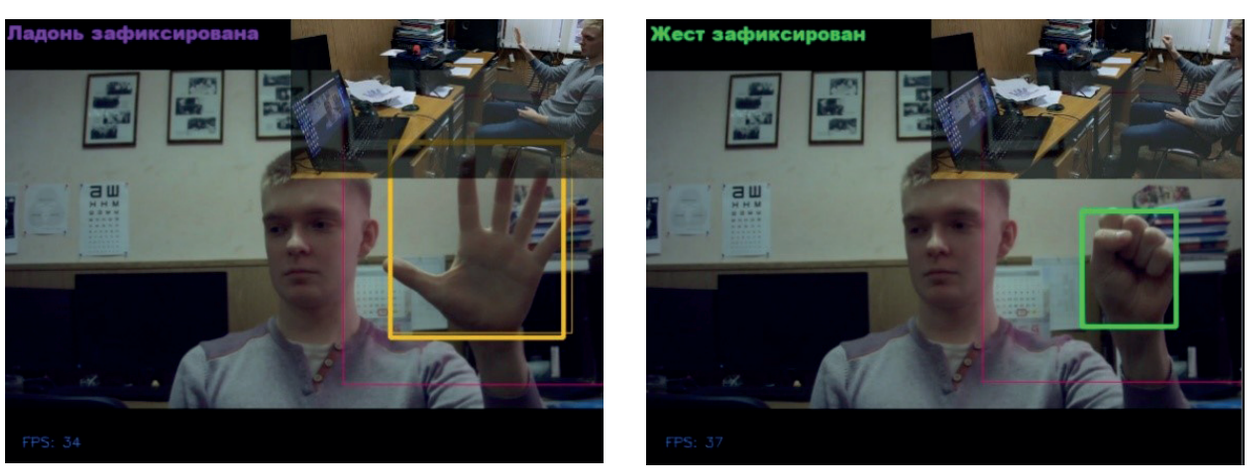
Недостатками данного метода являются меньший процент правильного детектирования элементов руки без перчатки. Но при этом он демонстрирует лучшую производительность и точность при определении руки в перчатке.
Указанные недостатки могут быть сокращены за счёт использования возможностей искусственного интеллекта.

## IV. Предлагаемый нейросетевой алгоритм

Для детектирования динамических жестов оператора предлагается использовать нейронную сеть семейства SSD (SingleShotDetector)[5] [39]. Ввиду ограничений, накладываемых работой в реальном масштабе времени, а также специальным вычислителем (одноплатный компьютер), была выбрана архитектура MobileNetSSD. На сегодняшний день существует 3 поколения данных архитектур. Первая версия появилась в 2016 году.
Входе исследований влияния архитектуры MobileNetSSDна точность детектирования жеста был сделан вывод о том, что оптимизированные слои во второй версии архитектуры негативно сказываются на производительности (без использования tf-lite), которая не может быть эффективно использована на одноплатных компьютерах без GPU или NPU модуля. Для повышения быстродействия и экономии ресурсов было решено применять MobileNetV1 с размером входного изображения 300 x 300 пикселей. Такой подход стал основой нашего проекта распознавания жестов. На рисунке 5 представлена составная часть выбранной архитектуры.




Р и с. 5. Архитектура MobileNetV1 300
F i g. 5. Architecture MobileNetV1 300




Р и с. 6. Демонстрация работы нейро-алгоритма
F i g. 6. The neuro-algorithm demonstration

Для обучения и верификации сети был составлен DataSet из нескольких наборов данных. В него вошли кадры из видео файлов, полученных разными камерами для группы студентов с разным цветом кожи, обучающихся в университете. В общей сложности объем DataSet достиг примерно пяти тысяч изображений. Данные во время обучения из DataSetподавались случайными пачками. Каждая пачка содержала по одному изображению каждого класса.

[5] Обработка изображений в авиационных системах технического зрения / Под ред. Л. Н. Костяшкина, М. Б. Никифорова. М.: Физматлит, 2018.

## Результаты

После обучения MobileNetV1 устойчиво распознавала два выделенных класса (ладонь, кулак) с вероятностью до 85% (процент высоты объекта составил ~25% от высоты изображения) на расстоянии не более полутора метров (рисунок 6).
Кроме дальности расположения объекта от камеры, существует и ряд других проблем, которые могут повлиять на точность работы нейронной сети. Одной из таких проблем являются ложные срабатывания на лицо человека. Это происходит из-за сложности извлечения признаков ввиду недостаточной глубины нейронной сети.
Результаты тестирования размещены в таблице 2.

Таблица 2. **Сравнение предложенного нейросетевого метода с аналитическим**

Table 2. **Comparison of the proposed neural network method with the analytical method**

| | **Нейросетевой алгоритм** | **Аналитический алгоритм** |
|---|---|---|
| Потребление памяти | 30 Мб памяти | Менее 1Мб памяти |
| Кол-во кадров на IntelAtom Z8500 | 5 FPS | 30 FPS |
| Процент правильного детектирования ладони (ладонь занимает 60% от высоты кадра) | 98% | 90% |
| Процент правильного детектирования ладони в перчатке (ладонь занимает 60% от высоты кадра) | 70% | 83% |
| Процент правильного детектирования ладони (ладонь занимает 40% от высоты кадра) | 94% | 83% |
| Процент правильного детектирования ладони в перчатке (ладонь занимает 40% от высоты кадра) | 77% | 81% |
| Процент правильного детектирования жеста «кулак» (занимает 20% от высоты кадра) | 85 % | 24% |
| Процент ошибочного детектирования «Кулак» при размере объекта 60% от высоты кадра | 5% | 6% |
| Количество детектируемых жестов | Обучается сетью | 5-6 разных жестов |

При использовании одноплатного компьютера без специального вычислителя или дополнительного оборудования для построения 3D-модели руки получить большую точность и скорость работы не предоставляется возможным.

## V. Заключение

Предложенные алгоритмы можно интегрировать в SoC. Но для работы нейросетевого алгоритма в режиме реального времени требуется интегрированный модуль NPU. Встроенной видеокарты в кристаллы большинства одноплатных компьютеров недостаточно для обеспечения требуемой производительности нейросетевого алгоритма. Точность нейросетевого метода выше, чем у аналитического. Аналитический метод позволяет детектировать жесты руки в перчатке лучше, чем существующие решения. Также стоит отметить, что нейросетевой алгоритм работает лучше в сложных условиях (перепады яркости, подвижный фон).
Ближайшими перспективами развития проекта является увеличение скорости работы нейросетевого алгоритма без применения специальных вычислителей.

## References

[1] Cho Y., Lee A., Park J., Ko B., Kim N. Enhancement of gesture recognition for contactless interface using a personalized classifier in the operating room. *Computer Methods and Programs in Biomedicine*. 2018; 161:39-44. (In Eng.) DOI: https://doi.org/10.1016/j.cmpb.2018.04.003

[2] Wang P. et al. Large-scale Continuous Gesture Recognition Using Convolutional Neural Networks. *arXiv*:1608.06338, 2016. Available at: https://arxiv.org/abs/1608.06338 (accessed 21.08.2020). (In Eng.)

[3] Nahapetyan V.E., Khachumov V.M. Automatic transformation of Russian manual-alphabet gestures into textual form. *Scientific and Technical Information Processing*. 2014; 41(5):302-308. (In Eng.) DOI: https://doi.org/10.3103/S0147688214050037

[4] Kato M., Chen Y.-W., Xu G. Articulated hand tracking by PCA-ICA approach. In; *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*. Southampton, UK; 2006. p. 329-334. (In Eng.) DOI: https://doi.org/10.1109/FGR.2006.21

[5] Wachs J.P., Kölsch M., Stern H., Edan Y. Vision-based hand-gesture applications. *Communications of the ACM*. 2011; 54(2):60-71. (In Eng.) DOI: https://doi.org/10.1145/1897816.1897838

[6] Szeliski R. Computer Vision. Texts in Computer Science. *Texts in Computer Science*. Springer, London; 2011. (In Eng.) DOI: https://doi.org/10.1007/978-1-84882-935-0

[7] Forsyth D.A., Ponce J. Computer Vision: A Modern Approach. 2nd ed. Prentice Hall; US; 2002. (In Eng.)

[8] Aggarwal J.K., Cai Q. Human Motion Analysis: A Review. *Computer Vision and Image Understanding*. 1999; 73(3):428-440. (In Eng.) DOI: https://doi.org/10.1006/cviu.1998.0744

[9] Rogalla O., Ehrenmann M., Zöllner R., Becher R., Dillmann R. Using gesture and speech control for commanding a robot assistant. In: *Proceedings of the 11th IEEE International Workshop on Robot and Human Interactive Communication*. Berlin, Germany; 2002. p. 454-459. (In Eng.) DOI: https://doi.org/10.1109/ROMAN.2002.1045664

[10] Schlömer T., Poppinga B., Henze N., Boll S. Gesture recognition with a Wii controller. In: *Proceedings of the 2nd international conference on Tangible and embedded interaction (TEI '08).* Association for Computing Machinery, New York, NY, USA; 2008. p. 11-14. (In Eng.) DOI: https://doi.org/10.1145/1347390.1347395

[11] Cheok M.J., Omar Z., Jaward M.H. A review of hand gesture and sign language recognition techniques. *International Journal of Machine Learning and Cybernetics*. 2019; 10(1):131-153. (In Eng.) DOI: https://doi.org/10.1007/s13042-017-0705-5

[12] Chethana N.S., Divyaprabha, Kurian M.Z. Design and Implementation of Static Hand Gesture Recognition System for Device Control. In: Shetty N., Prasad N., Nalini N. (ed.) *Emerging Research in Computing, Information, Communication and Applications*. 2016; 3:589-596. Springer, Singapore. (In Eng.) DOI: https://doi.org/10.1007/978-981-10-0287-8_54

[13] Lugaresi C. et al. Media Pipe: A Framework for Building Perception Pipelines. *arXiv*:1906.08172v1, 2019. Available at: https://arxiv.org/abs/1906.08172 (accessed 21.08.2020). (In Eng.)

[14] Mestetskiy L., Bakina I., Kurakin A. Hand Geometry Analysis by Continuous Skeletons. In: Kamel M., Campilho A. (ed.) Image Analysis and Recognition. ICIAR 2011. Lecture Notes in Computer Science. 2011; 6754:130-139. Springer, Berlin, Heidelberg. (In Eng.) DOI: https://doi.org/10.1007/978-3-642-21596-4_14

[15] Mitra S., Acharya T. Gesture Recognition: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. 2007; 37(3):311-324. (In Eng.) DOI: https://doi.org/10.1109/TSMCC.2007.893280

[16] Holte M.B., Moeslund T.B., Fihl P. Fusion of range and intensity information for view invariant gesture recognition. In: *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Anchorage, AK, USA; 2008. p. 1-7. (In Eng.) DOI: https://doi.org/10.1109/CVPRW.2008.4563161

[17] Ren Z., Yuan J., Zhang Z. Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In: *Proceedings of the 19th ACM international conference on Multimedia (MM '11)*. Association for Computing Machinery, New York, NY, USA; 2011. p. 1093-1096. (In Eng.) DOI: https://doi.org/10.1145/2072298.2071946

[18] Liao B., Li J., Ju Z., Ouyang G. Hand Gesture Recognition with Generalized Hough Transform and DC-CNN Using Realsense. In: *2018 Eighth International Conference on Information Science and Technology (ICIST)*. Cordoba, Granada, and Seville, Spain; 2018. p. 84-90. (In Eng.) DOI: https://doi.org/10.1109/ICIST.2018.8426125

[19] Mantecón T., del-Blanco C.R., Jaureguizar F., García N. Hand Gesture Recognition Using Infrared Imagery Provided by Leap Motion Controller. In: Blanc-Talon J., Distante C., Philips W., Popescu D., Scheunders P. (ed.) Advanced Concepts for Intelligent Vision Systems. ACIVS 2016. *Lecture Notes in Computer Science*. 2016; 10016:47-57. Springer, Cham. (In Eng.) DOI: https://doi.org/10.1007/978-3-319-48680-2_5

[20] McCannon B.C. Rock Paper Scissors. *Journal of Economics*. 2007; 92(1):67-88. (In Eng.) DOI: https://doi.org/10.1007/s00712-007-0263-5

[21] Garg P., Aggarwal N., Sofat S. Vision Based Hand Gesture Recognition. *International Journal of Computer and Information Engineering*. 2009; 3(1):972-977. Available at: https://publications.waset.org/10237/pdf (accessed 21.08.2020). (In Eng.)

[22] Wu Y., Huang T.S. View-independent recognition of hand postures. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*. Hilton Head, SC, USA. 2000; 2:88-94. (In Eng.) DOI: https://doi.org/10.1109/CVPR.2000.854749

[23] Huang C.-L., Jeng S. A model-based hand gesture recognition system. *Machine Vision and Applications*. 2001; 12(5):243-258. (In Eng.) DOI: https://doi.org/10.1007/s001380050144

[24] Wu D., Zhu F., Shao L. One shot learning gesture recognition from RGBD images. In: *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Providence, RI, USA; 2012. p. 7-12. (In Eng.) DOI: https://doi.org/10.1109/CVPRW.2012.6239179

[25] Keskin C., Kiraç F., Kara Y.E., Akarun L. Randomized decision forests for static and dynamic hand shape classification. In: *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Providence, RI, USA; 2012. p. 31-36. (In Eng.) DOI: https://doi.org/10.1109/CVPRW.2012.6239183

[26] Dominio F., Donadeo M., Zanuttigh P. Combining multiple depth-based descriptors for hand gesture recognition. *Pattern Recognition Letters*. 2014; 50:101-111. (In Eng.) DOI: https://doi.org/10.1016/j.patrec.2013.10.010

[27] Ren Z., Meng J., Yuan J., Zhang Z. Robust hand gesture recognition with kinect sensor. In: *Proceedings of the 19th ACM international conference on Multimedia (MM '11)*. Association for Computing Machinery, New York, NY, USA; 2011. p. 759-760. (In Eng.) DOI: https://doi.org/10.1145/2072298.2072443

[28] Yuan Q., Sclaroff S., Athitsos V. Automatic 2D Hand Tracking in Video Sequences. In: *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)*. Breckenridge, CO, USA; 2005. vol. 1, p. 250-256. (In Eng.) DOI: https://doi.org/10.1109/ACVMOT.2005.27

[29] van den Bergh M. et al. Real-time 3D hand gesture interaction with a robot for understanding directions from humans. In: *2011 RO-MAN*. Atlanta, GA, USA; 2011. p. 357-362. (In Eng.) DOI: https://doi.org/10.1109/ROMAN.2011.6005195

[30] Munoz-Salinas R., Medina-Carnicer R., Madrid-Cuevas F.J., Carmona-Poyato A. Depth silhouettes for gesture recognition. *Pattern Recognition Letters*. 2008; 29(3):319-329. (In Eng.) DOI: https://doi.org/10.1016/j.patrec.2007.10.011

[31] Lucas B.D., Kanade T. An Iterative Image Registration Technique with an Application to Stereo Vision. In: *Proceedings of the 7th international joint conference on Artificial intelligence — Vol. (IJCAI'81)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA; 1981. p. 674-679. (In Eng.)

[32] Stenger B., Mendonça P.R.S., Cipolla R. Model-Based 3D Tracking of an Articulated Hand. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. Kauai, HI, USA; 2001. p. II-II. (In Eng.) DOI: https://doi.org/10.1109/CVPR.2001.990976

[33] Yacoob Y., Davis L.S. Recognizing human facial expressions from long image sequences using optical flow. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1996; 18(6):636-642. (In Eng.) DOI: https://doi.org/10.1109/34.506414

[34] Zeng Z., Gong Q., Zhang J. CNN Model Design of Gesture Recognition Based on Tensorflow Framework. In: *2019 IEEE*

*3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. Chengdu, China; 2019. p. 1062-1067. (In Eng.) DOI: https://doi.org/10.1109/ITNEC.2019.8729185

[35] Pisharady P.K., Saerbeck M. Recent methods and databases in vision-based hand gesture recognition: A review. *Computer Vision and Image Understanding*. 2015; 141:152-165. (In Eng.) DOI: https://doi.org/10.1016/j.cviu.2015.08.004

[36] Han X.H., Chen Y.W., Nakao Z. Robust Edge Detection by Independent Component Analysis in Noisy Images. *IEICE TRANSACTIONS on Information and Systems*. 2004; E87-D(9):2204-2211. (In Eng.)

[37] Shi J., Tomasi C. Good Features to Track. In: *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA; 1994. p. 593-600. (In Eng.) DOI: https://doi.org/10.1109/CVPR.1994.323794

[38] Bansal M., Kumar Mu., Kumar Ma., Kumar K. An efficient technique for object recognition using Shi-Tomasi corner detection algorithm. *Soft Computing*. 2012; 25(6):4423-4432. (In Eng.) DOI: https://doi.org/10.1007/s00500-020-05453-y

[39] Liu P., Li X., Cui H., Li S., Yuan Y. Hand Gesture Recognition Based on Single-Shot Multibox Detector Deep Learning. *Mobile Information Systems*. 2019; 2019:3410348. (In Eng.) DOI: https://doi.org/10.1155/2019/3410348

*Поступила 21.08.2020; одобрена после рецензирования 15.10.2020; принята к публикации 25.11.2020.*
*Submitted 21.08.2020; approved after reviewing 15.10.2020; accepted for publication 25.11.2020.*

Об авторах:

**Муратов Евгений Рашитович**, доцент кафедры электронных вычислительных машин, ФГБОУ «Рязанский государственный радиотехнический университет имени В.Ф. Уткина» (390005, Российская Федерация, г. Рязань, ул. Гагарина, д. 59/1), кандидат технических наук, доцент, ORCID: http://orcid.org/0000-0002-1664-3954, myratov_er@mail.ru

**Никифоров Михаил Борисович**, директор НОЦ «СпецЭВМ», заместитель заведующего кафедрой электронных вычислительных машин, ФГБОУ «Рязанский государственный радиотехнический университет имени В.Ф. Уткина» (390005, Российская Федерация, г. Рязань, ул. Гагарина, д. 59/1), кандидат технических наук, доцент, член-корреспондент Академии информатизации образования, ORCID: http://orcid.org/0000-0002- 4796-0776, nikiforov.m.b@mail.ru

**Скачков Артем Михайлович**, магистрант кафедры электронных вычислительных машин, ФГБОУ «Рязанский государственный радиотехнический университет имени В.Ф. Уткина» (390005, Российская Федерация, г. Рязань, ул. Гагарина, д. 59/1), ORCID: http://orcid.org/0000-0001-7902-7668, art.skachkov10@gmail.com

*Все авторы прочитали и одобрили окончательный вариант рукописи.*

About the authors:

**Yevgeniy R. Muratov**, Associate Professor of the Department of Electronic Computers Machines, Ryazan State Radio Engineering University named after V.F. Utkin (59/1 Gagarin St., Ryazan 390005, Russian Federation), Ph.D. (Engineering), Associate Professor, ORCID: http://orcid.org/0000-0002-1664-3954, myratov_er@mail.ru

**Mikhail B. Nikiforov**, Director of the SEC "SpecEVM", vice-head of the Department of Electronic Computing Machines, Ryazan State Radio Engineering University named after V.F. Utkin (59/1 Gagarin St., Ryazan 390005, Russian Federation), Ph.D. (Engineering), Associate Professor, Corresponding member of the Academy of Informatization of Education, ORCID: http://orcid.org/0000-0002- 4796-0776, nikiforov.m.b@mail.ru

**Artem M. Skachkov**, Master's student of the Department of Electronic Computers Machines, Ryazan State Radio Engineering University named after V.F. Utkin (59/1 Gagarin St., Ryazan 390005, Russian Federation), ORCID: http://orcid.org/0000-0001-7902-7668, art.skachkov10@gmail.com

*All authors have read and approved the final manuscript.*