

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPUTER ENGINEERING
AND CONTROL

УДК 004.932.2

DOI: 10.17212/1814-1196-2020-1-25-40

Распознавание жестов рук в режиме реального времени^{*}

Д.А. БУЛЫГИН^а, Т.Е. МАМОНОВА^б

634034, РФ, г. Томск, пр. Ленина, 30, Национальный исследовательский Томский политехнический университет

^а dab28@tpu.ru ^б stepte@tpu.ru

В настоящее время всё больше исследований направлено на решение задач с применением компьютерного зрения и искусственного интеллекта. Наиболее частыми являются решения и подходы с использованием распознавания жестов на основании инфракрасных сенсоров или нейронных сетей.

Актуальность рассматриваемой тематики обусловлена возможностью применения предлагаемого подхода для управления работой объектов без тактильного контакта и голосовой идентификации команд, а также своей простотой с точки зрения конечного пользователя.

В настоящей работе проанализированы существующие способы распознавания жестов. Рассмотрены методы и подходы, а также их реализация, исследованы преимущества и недостатки рассмотренных методов. На их основе составлена таблица с тезисной информацией и предложена собственная архитектура сверточной нейронной сети для решения классификации жестов. Проведена оценка точности работы сети. На основе полученных данных проведен двухфакторный анализ зависимости сложности жеста, его дальности и точности полученного алгоритма.

По полученной зависимости построены графики изменения точности работы сверточной нейронной сети. Проанализирован характер изменения точности для различных факторов.

Ключевые слова: распознавание жестов, компьютерное зрение, инфракрасные датчики, сверточные нейронные сети, обучение, support vector machine, классификация, Keras, Tensorflow

ВВЕДЕНИЕ

Распознавание жестов играет важную роль во взаимодействии человека с машиной из-за его естественного и дружественного семантического выражения. Для использования этой технологии машины должны быстро и точно их определять, чтобы пользователи чувствовали себя комфортно и были готовы взаимодействовать с машинами. Распознавание жестов остается слож-

^{*} Статья получена 03 декабря 2019 г.

ной задачей из-за их разнообразия, сходства форм и сложности сценариев применения.

Задача распознавания жестов имеет различные решения в научных работах. Существующие подходы обычно можно разделить на две категории:

- 1) распознавание на основе носимых устройств,
- 2) распознавание на основе компьютерного зрения.

Одной из первых технологий для решения задачи распознавания жестов были специальные перчатки [1, 2]. Эта технология используется для сбора данных и их дальнейшей передачи. Данными являются движения рук, такие как скорость и углы поворота. Эти данные приходят на компьютер, и компьютер выполняет распознавание с помощью специального алгоритма. Например, Такахаша и Кишино разработали перчатку данных, способную распознавать 46 видов жестов [3]. Yangsheng et al. использовали перчатки Saibo, чтобы точно идентифицировать 14 различных жестов и затем управлять роботом [4]. Использование специальных перчаток для решения задачи распознавания жестов не получило широкого распространения из-за неудобства использования.

В последнее время для решения задач распознавания жестов рук всё чаще появляются методы и решения, не требующие специальных перчаток. Далее выполнен анализ существующих способов распознавания жестов при помощи нейронных сетей или инфракрасных сенсоров.

1. ИНФРАКРАСНЫЕ СЕНСОРЫ НА ОСНОВЕ INTEL REALSENSE

Французскими учеными Квентином Смедтом, Хэземом Ванноусом и Джин-Филиппе Вандеборре была предложена идея распознавания жестов при помощи данных о скелете кисти в трехмерном формате [5].

В настоящей работе используется устройство Intel RealSense. В его основе лежат инфракрасные сенсоры, при помощи которых можно получить информацию не только об изображении руки, но и о ее положении в пространстве. Эта информация позволяет судить о том, насколько далеко находится рука от инфракрасного сенсора, т. е. о глубине.

На рис. 1 изображена информация о глубине и скелете кисти, полученная при помощи Intel RealSense. Сенсоры возвращают 22 соединения, по четыре на каждый палец, одно в центре ладони и одно на запястье.

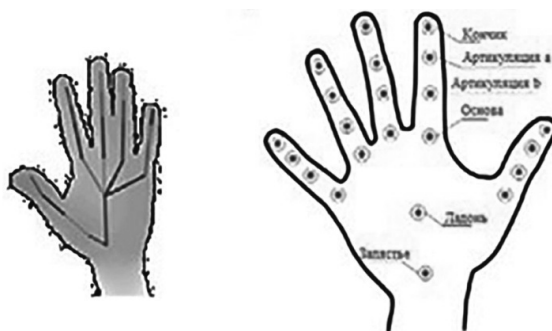


Рис. 1. Информация, полученная при помощи Intel RealSense

Fig. 1. Information obtained with Intel RealSense

Используя трехмерные данные скелета руки, изображенные на рис. 1, динамический жест можно рассматривать как временной ряд скелетов руки. Он описывает движение и форму руки, обозначающую жест. Для каждого кадра последовательность положения скелета руки представлена тремя координатами.

Скелеты рук, возвращаемые датчиками, состоят из трехмерных координат суставов рук, представленных в системе координат камеры. Поэтому они меняются в зависимости от поворота и перемещения руки относительно камеры.

Затем дескрипторы описывают форму руки и изменение движения внутри последовательности положений руки, но они не учитывают динамический характер жеста. Чтобы добавить временную метку, используется «Временная пирамида» (Temporal Pyramid).

Для классификации жестов используется контролируемый классификатор SVM с линейным ядром, поскольку он легко справляется с многомерным представлением.

Полная модель распознавания жестов при помощи устройства Intel RealSense изображена на рис. 2.

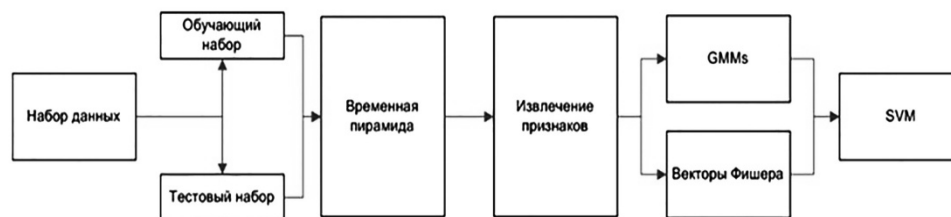


Рис. 2. Полная модель распознавания жестов

Fig. 2. Pipeline of gesture recognition approach

ДИНАМИЧЕСКОЕ РАСПОЗНАВАНИЕ ЖЕСТОВ

Динамическое распознавание жестов по направленным импульсным нейронным сетям (PCNN) для взаимодействия человека с роботом в режиме реального времени предложено китайскими учеными и раскрыто в работе [6].

Здесь предлагается модифицированный алгоритм PCNN, называемый прямым PCNN (DPCNN), для распознавания жестов в реальном времени без большого количества обучающих данных. Трехмерные координаты ключевых каркасных точек тела человека извлекаются в качестве входного объекта в виде последовательности из данных, захваченных Kinect, показанных на рис. 3. Чтобы распознать жест ввода, сходство с ним вычисляется из матрицы объектов, созданной из элементов ввода и заданного шаблона последовательностей признаков.

Для эффективного распознавания жестов матрица признаков преобразуется в неориентированный граф. Сходство проблемы распознавания жестов преобразуется в задачу кратчайшего пути. Чтобы быстро и точно решить кратчайший путь, DPCNN выбирает направление движения, стимулируя соседние нейроны различными возбуждениями, и уменьшает эффект бесполезных нейронов. Кроме того, чтобы сократить время распознавания, введен

способ раннего распознавания жестов, основанный на адаптивном окне, для распознавания неполных жестов. Предложенный метод раннего распознавания нацелен на повышение эффективности распознавания жестов в реальном времени и взаимодействие человека и робота.

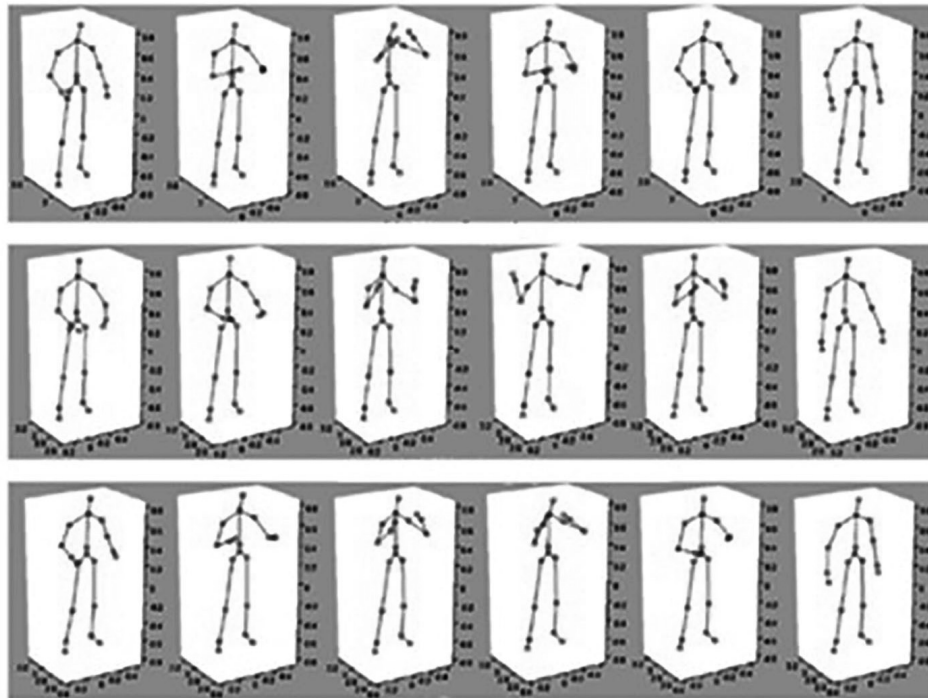


Рис. 3. Последовательность трехмерных данных

Fig. 3. The sequences of data

Полная модель в виде последовательных блоков представлена на рис. 4.



Рис. 4. Полная модель алгоритма

Fig. 4. The flowchart of the dynamic gesture recognition

Во время потока динамического распознавания жестов наиболее важным шагом является быстрое и точное вычисление кратчайшего пути нейронной сетью с импульсной связью (PCNN). Тип сигнала и механизм обработки PCNN подобен физиологической основе в системе зрительных нейронов человека. PCNN выбирается для задачи с кратчайшим путем и вычисляет сходства для динамического распознавания жестов.

Структуру PCNN можно рассматривать как граф, обозначающий множество узлов (нейронов). При работе с задачей кратчайшего пути алгоритм PCNN активирует соседние нейроны совершенно недетерминированным способом. Таким образом, вычислительная сложность PCNN зависит от масштаба сети. По мере увеличения масштаба сети время и затраты на вычисления также увеличиваются. Из-за этого недостатка PCNN может занимать много времени при распознавании жестов в реальном времени, несмотря на его высокую точность.

Разработанный алгоритм DPCNN выбирает направление потока, стимулируя соседние нейроны различными возбуждениями, и рассчитывает кратчайший путь, который рассматривается как сходство между последовательностями признаков тестов и образцов. Кроме того, алгоритм раннего распознавания, основанный на адаптивном окне, введен для лучшего взаимодействия между человеком и роботами.

РАСПОЗНАВАНИЕ ЖЕСТОВ В РЕАЛЬНОМ ВРЕМЕНИ НА ОСНОВЕ СЕТИ ПЕРЕКАЛИБРОВКИ ФУНКЦИЙ С МНОГОМАСШТАБНОЙ ИНФОРМАЦИЕЙ

Последняя работа также была представлена китайскими учеными [7]. Они использовали сверточные нейронные сети для решения задачи распознавания и классификации жестов.

Существует две проблемы в процессе распознавания жеста, расположенного на большом расстоянии от камеры. Во-первых, жесты с различным соотношением размеров трудно идентифицировать; во-вторых, существует разная информация между признаками низкого уровня и признаками высокого уровня. Несмотря на то что низкоуровневые признаки с высоким разрешением содержат больше деталей и информацию о позиции в кадре, это не способствует выявлению нужных признаков разных размеров при обнаружении. И, напротив, высокоуровневые признаки больше подходят для классификации по категориям, но имеют более низкое восприятие деталей из-за их более низкого разрешения.

В некоторых сетях, таких как SegNet и Unet, точность и надежность сети улучшаются путем объединения нескольких наборов низкоуровневых признаков и остальных признаков, но эти методы неэффективны для практического использования.

В основе настоящей работы лежит объединение признаков в разных масштабах. Чтобы лучше извлечь контекстную информацию разных масштабов извлекается информация о характеристиках разных масштабов при помощи сверточного ядра с размером шага 2 и размером 3×3 и 5×5 соответственно.

Полная структура сверточной нейронной сети изображена на рис. 5.

Модуль FPA используется для объединения различных масштабов контекстной информации для объединения локальной и глобальной информации. Кроме того, применение ядра свертки эффективно уменьшает количество параметров в тренировочном процессе и повышает скорость

распознавания жестов. Для обучения сети для решения задачи взаимодействия человека с машиной было собрано 3289 изображений, содержащих 16 жестов. Затем был использован метод изменения насыщенности оттенка и вращения изображений, чтобы увеличить размер набора данных, тем самым увеличив устойчивость нейронной сети. По сравнению с популярными в настоящее время сверточными сетями метод может распознавать жесты в режиме реального времени, обеспечивая при этом высокую точность.

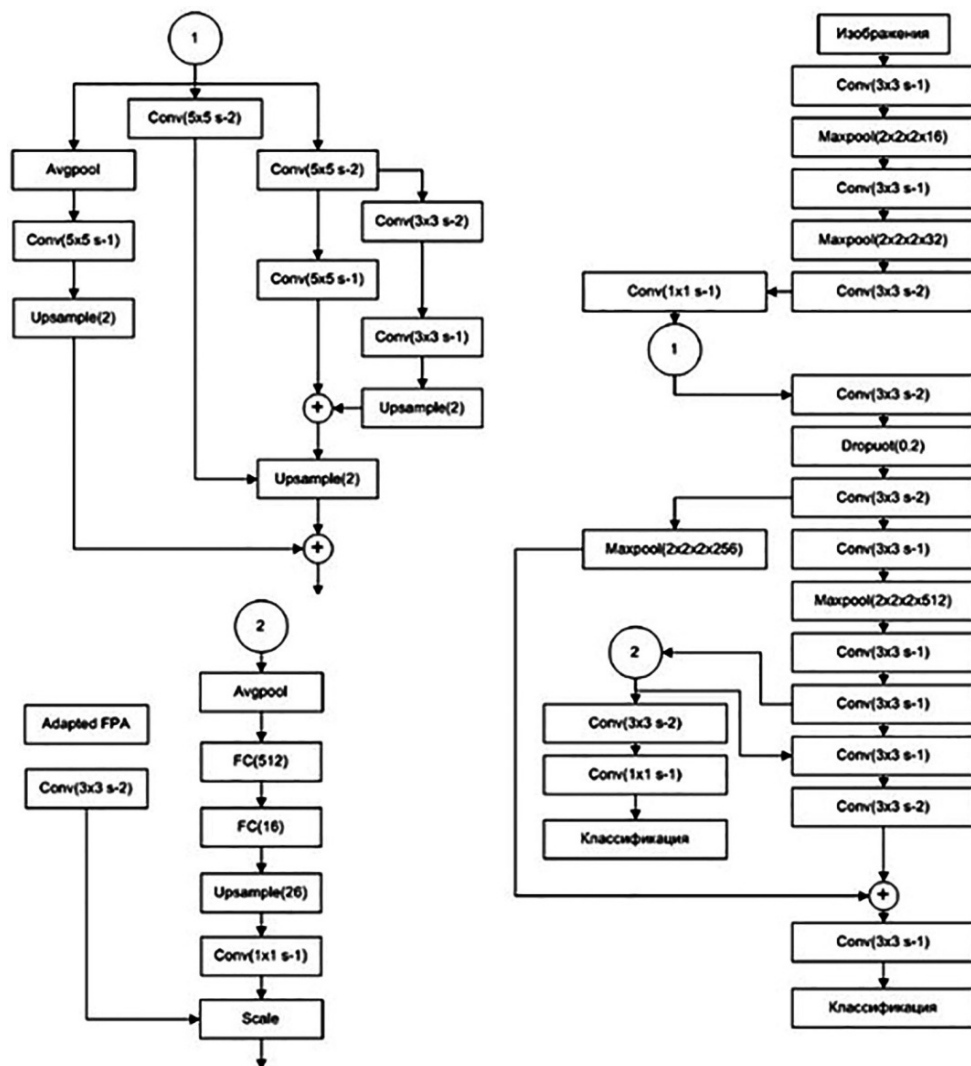


Рис. 5. Общая структура нейронной сети

Fig. 5. The gesture recognition network structure diagram

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПРИВЕДЕННЫХ МЕТОДОВ РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ ЖЕСТОВ

Краткое описание рассмотренных алгоритмов для решения задачи распознавания жестов рук приведено в табл. 1.

Таблица 1

Table 1

Краткое описание изученных алгоритмов

A brief description of the algorithms

№ п/п	Название	Оборудование	Входные величины	Алгоритм	Преимущества	Недостатки
1	Распознавание жестов при помощи инфракрасных сенсоров на основе Intel RealSense	Устройство Intel RealSense, вычислительная машина	Трехмерные данные о скелете руки	Использование инфракрасных сенсоров, получение соединений с последующей классификацией	Отсутствие влияния света на работу алгоритма, высокая точность распознавания	Необходимость наличия дополнительного устройства Intel RealSense, небольшая дальность работы, сложность алгоритма
2	Динамическое распознавание жестов по направленным импульсным нейронным сетям	Устройство Microsoft Kinect, вычислительная машина	Трехмерные данные о ключевых точках корпуса человека	Преобразование задачи распознавания в задачу кратчайшего пути при помощи PCNN	Отсутствие влияния света на работу алгоритма, скорость работы алгоритма	Сложность реализации, большие вычислительные затраты, необходимость наличия дополнительного устройства Microsoft Kinect
3	Распознавание жестов в реальном времени на основе сети перекалибровки функций с многомасштабной информацией	Камера, вычислительная машина	Изображение, подающееся на вход сверточного слоя	Обработка изображения сверточными слоями для вычленения признаков различных уровней	Высокая точность классификации жестов, хорошая дальность работы	Большая архитектура нейронной сети, невысокая скорость работы

Исходя из табл. 1 можно видеть, что рассмотренные методы обладают высокой точностью распознавания и классификации, но алгоритмы сложны и не всегда могут работать с высокой скоростью из-за высокой сложности.

2. РАСПОЗНАВАНИЕ ЖЕСТОВ ПРИ ПОМОЩИ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ С НЕБОЛЬШОЙ АРХИТЕКТУРОЙ

Для увеличения быстродействия предложен алгоритм, в основе которого находится сверточная нейронная сеть с небольшой архитектурой.

Основой сверточной нейронной сети являются слои свертки [5]. Каждый слой включает в себя фильтры для каждого канала. Они обрабатывают предыдущий слой по частям (путем суммирования матричных фрагментов). Все веса ядра свертки заранее неизвестны и изменяются в процессе обучения в зависимости от входных данных. В конце слоя свертки всегда стоит функция активации [8].

Для передачи информации с одного слоя на другой необходима функция активации. Она преобразует информацию (численные значения) со всех нейронов предыдущего слоя в определенное значение для нейрона текущего слоя. Выход зависит от функции активации и может быть как действительным, так и целым [9]. Значение выхода – это показатель того, насколько активировался нейрон текущего слоя.

Так как каждому фильтру свертки соответствует одна карта признаков, то это позволяет нейронной сети научиться выделять признаки независимо от их расположения во входном изображении.

Пуллинг можно истолковать так: если на предыдущей операции свертки были обнаружены некоторые признаки, то для дальнейшей обработки настолько подробное изображение уже не нужно, и оно уменьшается в размерности, т. е. уплотняется до менее подробного. К тому же фильтрация уже ненужных деталей уменьшает переобучение. Слой пуллинга, как правило, вставляется после слоя свертки перед слоем следующей свертки.

За счет слоя пуллинга сеть становится наиболее устойчивой к изменениям входного изображения, например, к его сдвигам. Также уменьшается размерность последующих слоев [10].

Полносвязный слой (многослойный перцептрон) – скрытый слой, соединенный со всеми нейронами предыдущего слоя. Последним слоем многослойного перцептрона является один или несколько нейронов, количество которых равно количеству классов. Проще говоря, на вход всей сверточной нейронной сети подается изображение, а на выходе сеть выдает класс, к которому это изображение относится [11].

Сверточные нейронные сети обеспечивают частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям. Общая топология изображена на рис. 6.

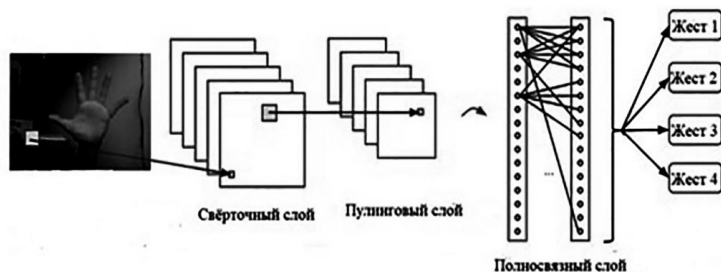


Рис. 6. Общая структура нейронной сети

Fig. 6. The gesture recognition network architecture

В результате была спроектирована сверточная нейронная сеть с параметрами, указанными в табл. 2.

Таблица 2

Table 2

Результаты проектирования нейронной сети

Results of the neural network design

Параметр		Характеристика
Количество сверточных слоев и слоев пуллинга		3
Размер фильтров	Первый слой	5×5
	Второй слой	3×3
Количество фильтров для каждого слоя		8, 12, 16
Вероятность Drop-out слоя		50 %
Количество нейронов в полносвязном слое		128

Данная сверточная нейронная сеть обучена на выборке из 12 000 изображений, соответствующих четырем классам. Количество эпох обучения равно трем. При большем количестве эпох сеть начинает «выучивать» данные с изображений и становится не способна работать с окружением, отличным от того, которое присутствует в обучающей выборке, т. е. происходит переобучение.

Общая точность обучения на тестовой выборке, полученная при помощи функций библиотеки для машинного обучения Keras [12], примерно равна 95 %. Результаты успешных определений жестов приведены на рис. 7.

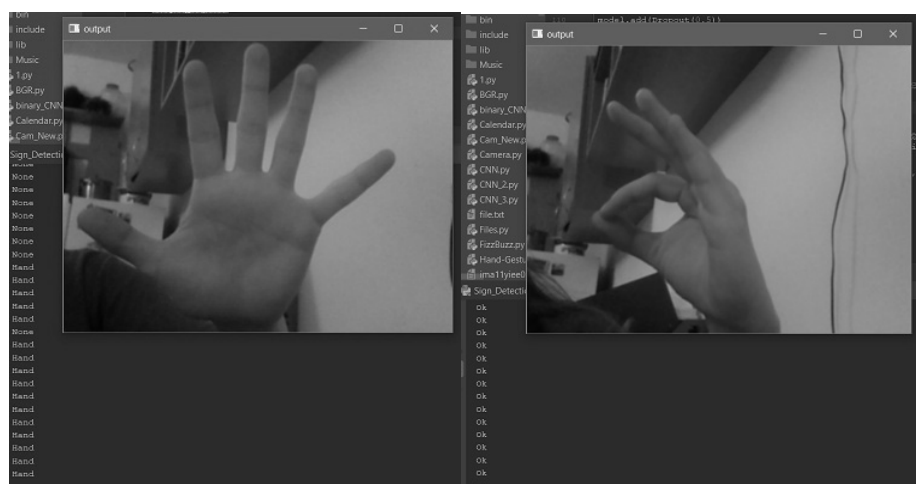


Рис. 7. Успешное обнаружение жестов нейронной сетью

Fig. 7. Successful gesture detection by neural network

Была исследована точность нейронной сети при изменении двух факторов: расстояния руки от веб-камеры и сложности жеста.

Точность измерялась из соотношения успешно распознанных кадров к общему числу кадров. Количество кадров было принято равным 50, так как экспериментально было получено, что точность работы сети не зависит от количества кадров.

Расстояние от веб-камеры было взято от 30 до 60 сантиметров.

Сложность жестов зависит от сложности распознавания, например, жест «Ок» можно легко спутать с жестом «Победа», поэтому сложность жеста равна трем. Диапазон сложности варьируется от одного до трех.

Для выявления зависимости точности работы нейросети от дальности и сложности жеста был произведен двухфакторный эксперимент [13]. Измеренные факторы в натуральном масштабе и точность отображены в табл. 3.

Таблица 3

Table 3

Результаты измерения

Measurement results

Номер опыта	Факторы в натуральном масштабе		Выходной параметр
	Z_1	Z_2	Y
1	30	1	94,2
2	60	1	89,7
3	30	3	91,6
4	60	3	83,9

Факторы из натурального масштаба были переведены в безразмерный масштаб при помощи формулы (1):

$$Z_i^* = \frac{z_i - Z_i^0}{\Delta z_i}; \quad (1)$$

$$Z_i^0 = \frac{z_i^{\max} + z_i^{\min}}{2}, \text{ где } i = 1, 2, 3, \dots, k; \quad (2)$$

$$\Delta z_i = \frac{z_i^{\max} - z_i^{\min}}{2}, \quad (3)$$

где z_i – значение фактора в натуральном масштабе; Z_i^* – значение фактора в безразмерном масштабе; z_i^{\max} – максимальное значение фактора; z_i^{\min} – минимальное значение фактора.

Была построена матрица планирования двухфакторного эксперимента. Матрица отображена в табл. 4.

Таблица 4

Table 4

Матрица планирования**The matrix of planning**

Номер опыта	X_0	X_1	X_2	Y
1	+1	−1	−1	94,2
2	+1	+1	−1	89,7
3	+1	−1	+1	91,6
4	+1	+1	+1	83,9

Линейные коэффициенты регрессии рассчитаны по формуле

$$b_i = \frac{1}{n} \sum_{i=1}^n y_i, \quad (4)$$

где b_i – значение коэффициента в уравнении регрессии; n – количество опытов; y_i – среднее значение точности.

Рассчитаны коэффициенты парного взаимодействия. Для этого составлена дополнительная таблица (табл. 5).

Таблица 5

Table 5

Расширенная матрица планирования полного факторного эксперимента**An augmented matrix of planning a full factorial experiment**

Номер опыта	X_0	X_1	X_2	X_1X_2	Y
1	+1	−1	−1	+1	94,2
2	+1	+1	−1	−1	89,7
3	+1	−1	+1	−1	91,6
4	+1	+1	+1	+1	83,9

Уравнение регрессии, полученное при проведении планирования эксперимента для двух факторов:

$$y(x_1, x_2) = 89,9 - 3,1x_1 - 2,1x_2 - 0,8x_1x_2.$$

Уравнение регрессии, состоящее из факторов натуральной величины, имеет вид

$$y(z_1, z_2) = 107,9 - 0,3z_1 - 4,3z_2 + 0,1z_1z_2.$$

Проверка коэффициентов, проведенная по критерию Стьюдента с использованием параллельных опытов, показала значимость всех коэффициентов полученного уравнения. Проверка адекватности уравнения осуществлялась с использованием критерия Фишера и показала, что математическая модель регрессии адекватна и может быть использована для исследований [14].

По полученному уравнению регрессии построена зависимость точности работы нейронной сети от дальности при сложности жеста, равной единице. Зависимость изображена на рис. 8.

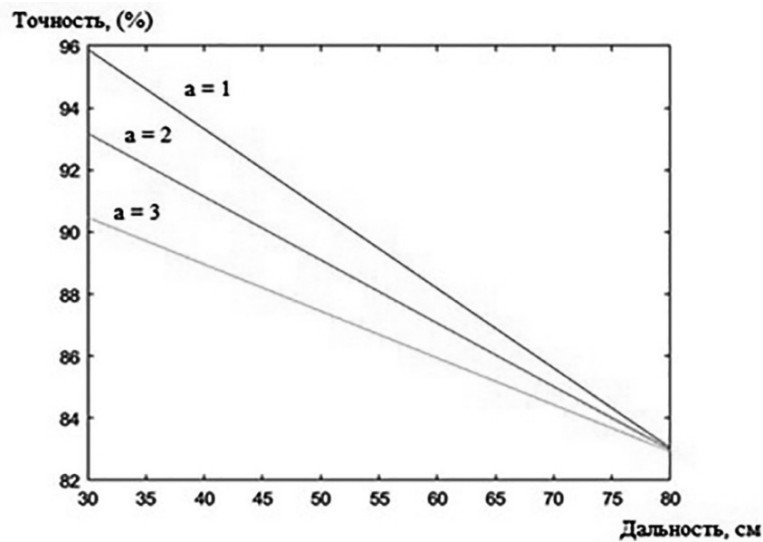


Рис. 8. Зависимость точности работы нейронной сети от дальности (a – сложность жеста)

Fig. 8. The dependence of the neural network accuracy on the range (a is a gesture difficulty)

Также по полученному уравнению регрессии построена зависимость точности работы нейронной сети от сложности жеста при расстоянии 60 см. Зависимость изображена на рис. 9. Наибольшую точность нейронная сеть имеет при сложности жеста, равной единице. При увеличении дальности точность сети при любой сложности жеста падает до 84 %.

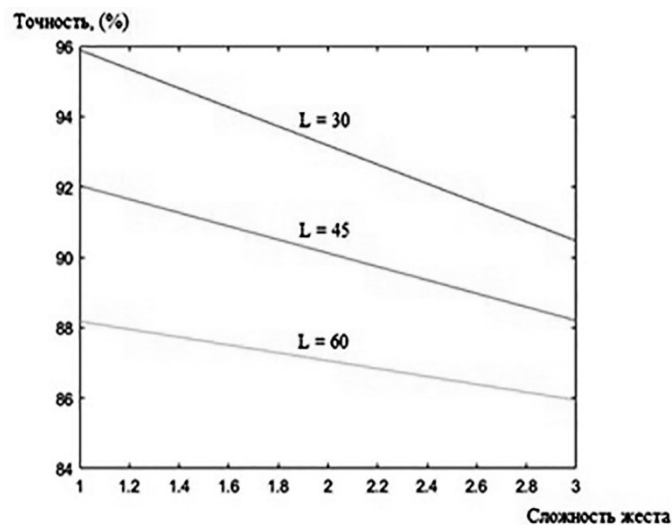


Рис. 9. Зависимость точности работы нейронной сети от сложности жеста (L – расстояние между жестом и камерой)

Fig. 9. The dependence of the neural network accuracy on the complexity of the gesture (L is the distance between the gesture and the camera)

Исходя из рис. 8 и 9 можно сделать вывод, что точность работы нейронной сети уменьшается при увеличении расстояния между рукой и камерой. Также точность падает при использовании сложных жестов по сравнению с простыми.

ЗАКЛЮЧЕНИЕ

Распознавание жестов важно для взаимодействия человека с машиной. Современные методы решения распознавания жестов имеют ряд недостатков, таких как низкая скорость распознавания, низкая скорость и низкая производительность при распознавании нескольких целей или целей на большом расстоянии в сложных условиях. Ввиду вышеупомянутых проблем был предложен подход распознавания жестов сверточной нейронной сетью с меньшей архитектурой.

Предложенная сверточная нейронная сеть имеет наименьшее время обучения, скорость работы и требует меньших затрат вычислительных мощностей. Точность распознавания уступает остальным приведенным методам в данной работе, но этот недостаток нивелируется скоростью работы и невысокими вычислительными затратами алгоритма.

В результате была разработана нейронная сеть, состоящая из трех сверточных слоев, распознающая три жеста или же их отсутствие, а также исследована ее работа и выявлены зависимости точности от дальности и сложности жестов. На небольшой дальности (30 см) точность быстрее уменьшается, но ее значения выше, чем при дальности 45 см и 60 см.

СПИСОК ЛИТЕРАТУРЫ

1. *Takahashi T., Kishino F.* Hand gesture coding based on experiments using a hand gesture interface device // ACM SIGCHI Bulletin. – 1991. – Vol. 23 (2). – P. 67–74.
2. *Lee C., Xu Y.* Online, interactive learning of gestures for human/robot interfaces // Proceedings of the IEEE International Conference on Robotics and Automation. – 2002. – Vol. 4. – P. 2982–2987.
3. *Smedt Q. De, Wannous H., Vandeborre J.-P.* Heterogeneous hand gesture recognition using 3D dynamic skeletal data // Computer Vision and Image Understanding. – 2019. – Vol. 181. – P. 60–72.
4. Dynamic gesture recognition by directional pulse coupled neural networks for human-robot interaction in real time / J. Dong, Z. Xia, W. Yan, Q. Zhao // Journal of Visual Communication and Image Representation. – 2019. – Vol. 63. – P. 102583.
5. Real-time gesture recognition based on feature recalibration network with multi-scale information / Z. Cao, X. Xu, B. Hu, M. Zhou, Q. Li // Neurocomputing. – 2019. – Vol. 347. – P. 119–130.
6. *Голиков И.* Сверточная нейронная сеть, часть 1: структура, топология, функции активации и обучающее множество. – URL: <https://habr.com/ru/post/348000/> (дата обращения: 12.12.2019).
7. Функции активации в нейронных сетях. – URL: <http://www.aiportal.ru/articles/neural-networks/activation-function.html> (дата обращения: 12.12.2019).
8. Библиотека Keras. Слой пуллинга. – URL: <https://keras.io/layers/pooling/> (дата обращения: 12.12.2019).
9. Библиотека Keras. Полносвязный слой. – URL: <https://keras.io/layers/core/> (дата обращения: 12.12.2019).
10. Библиотека Keras. Model class API. – URL: <https://keras.io/models/model/> (дата обращения: 12.12.2019).
11. Основы планирования эксперимента: методическое пособие / сост. К.М. Хамханов. – Улан-Удэ, 2001. – URL: <http://window.edu.ru/resource/438/18438/files/Mtduk8.pdf> (дата обращения: 12.12.2019).
12. Проверка адекватности регрессионной модели. – URL: <https://helpstat.ru/proverka-adekvatnosti-regressionnoj-modeli/> (дата обращения: 12.12.2019).

Булыгин Дмитрий Александрович, студент направления «Мехатроника и робототехника» Национального исследовательского Томского политехнического университета. Основное направление научных исследований – компьютерное зрение и машинное обучение. E-mail: 128dmitriy128@gmail.com.

Мамонова Татьяна Егоровна, кандидат технических наук, доцент отделения автоматизации и робототехники Национального исследовательского Томского политехнического университета. Основное направление научных исследований – моделирование течения жидкости в трубопроводе на основе гидродинамических процессов, искусственный интеллект и его применение, цифровизация процессов и систем, когнитивная визуализация данных, обработка изображений. Имеет 56 публикаций. E-mail: stepte@tpu.ru.

Bulygin Dmitriy A., student at the mechatronics and robotics department in TPU. His main research subjects are computer vision and machine learning. E-mail: 128dmitriy128@gmail.com.

Mamonova Tatyana E., PhD (Eng.), an associate professor at the TPU automation and robotics department. Her main research subjects are fluid flow modeling in a pipeline based on hydrodynamic processes, artificial intelligence and its application, digitalization of processes and systems, cognitive data visualization, and image processing. She is the author of 56 research papers. E-mail: stepte@tpu.ru.

DOI: 10.17212/1814-1196-2020-1-25-40

Recognition of hand gestures in real time*D.A. BULYGIN^a, T.E. MAMONOVA^b

Tomsk National Research Polytechnic University, 30 Lenin Street, Tomsk, 634034, Russian Federation

^a dab28@tpu.ru ^b stepte@tpu.ru**Abstract**

Gesture recognition plays an important role in the interaction of humans with machines because of their natural and friendly semantic expression. To use this technology, machines must quickly and accurately identify them so that users feel comfortable and are ready to interact with the machines. Gesture recognition remains a challenge due to the variety of gestures, the similarity of the forms of gestures and the complexity of the application scenarios.

Recently, more and more research in the field of computer vision is being conducted. The most common ones are approaches to recognizing gestures using infrared sensors or neural networks. In this paper, we analyze the existing solutions that allow us to solve the problem of gesture recognition using computer vision.

This topic is relevant since gesture control can be used in places where it is not possible to use voice commands. Also, this type of control is intuitive to humans.

In this paper, we analyze the existing methods for recognizing gestures using neural networks or infrared sensors. The advantages and disadvantages of the analyzed methods are investigated and the proprietary architecture of the convolutional neural network to classify gestures is proposed. Also, the accuracy of the network was evaluated depending on the distance between the camera and the hand, as well as depending on the complexity of the gesture.

Keywords: gesture recognition, computer vision, infrared sensors, convolutional networks, training, training data set, support vector machine, classification, Keras, Tensor flow

REFERENCES

1. Takahashi T., Kishino F. Hand gesture coding based on experiments using a hand gesture interface device. *ACM SIGCHI Bulletin*, 1991, vol. 23 (2), pp. 67–74.
2. Lee C., Xu Y. Online, interactive learning of gestures for human/robot inter-faces. *Proceedings of the IEEE International Conference on Robotics and Automation*, 2002, vol. 4, pp. 2982–2987.
3. Smedt Q. De, Wannous H., Vandeborre J.-P. Heterogeneous hand gesture recognition using 3D dynamic skeletal data. *Computer Vision and Image Understanding*, 2019, vol. 181, pp. 60–72.
4. Dong J., Xia Z., Yan W., Zhao Q. Dynamic gesture recognition by directional pulse coupled neural networks for human-robot interaction in real time. *Journal of Visual Communication and Image Representation*, 2019, vol. 63, p. 102583.
5. Cao Z., Xu X., Hu, Biao Z., Zhou M., Li Q. Real-time gesture recognition based on feature recalibration network with multi-scale information. *Neurocomputing*, 2019, vol. 347, pp. 119–130.
6. Golikov I. *Svertochnaya neironnaya set', chast' 1: struktura, topologiya, funktsii aktivatsii i obuchayushchee mnozhestvo* [Convolutional neural network, part 1: structure, topology, activation functions and training set]. Available at: <https://habr.com/ru/post/348000/> (accessed 12.12.2019).
7. *Funktsii aktivatsii v neironnykh setyakh* [Activation functions in neural networks]. Available at: <http://www.aiportal.ru/articles/neural-networks/activation-function.html> (accessed 12.12.2019).
8. Keras library. *Pooling layer*. Available at: <https://keras.io/layers/pooling/> (accessed 12.12.2019).
9. Keras library. *Fully connected layer*. Available at: <https://keras.io/layers/core/> (accessed 12.12.2019).

* Received 03 December 2019.

10. *Keras library. Model class API*. Available at: <https://keras.io/models/model/> (accessed 12.12.2019).

11. Khamkhanov K.M., comp. *Osnovy planirovaniya eksperimenta* [The basics of experiment planning]. Available at: <http://window.edu.ru/resource/438/18438/files/Mtdukm8.pdf> (accessed 12.12.2019).

12. *Proverka adekvatnosti regressionnoi modeli* [Verification of the adequacy of the regression model]. Available at: <https://helpstat.ru/proverka-adekvatnosti-regressionnoj-modeli/> (accessed 12.12.2019).

Для цитирования:

Булыгин Д.А., Мамонова Т.Е. Распознавание жестов рук в режиме реального времени // Научный вестник НГТУ. – 2020. – № 1 (78). – С. 25–40. – DOI: 10.17212/1814-1196-2020-1-25-40.

For citation:

Bulygin D.A., Mamonova T.E. Raspoznavanie zhestov ruk v rezhime real'nogo vremeni [Recognition of hand gestures in real time]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta = Science bulletin of the Novosibirsk state technical university*, 2020, no. 1 (78), pp. 25–40. DOI: 10.17212/1814-1196-2020-1-25-40.