ISSN 2782-2001
Системы анализа
и обработки данных
том 83, № 3, 2021, с. 53–74

http://journals.nstu.ru/vestnik
Analysis and data
processing systems
Vol. 83, No. 3, 2021, pp. 53–74

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPPUTER ENGINEERING
AND CONTROL

УДК 004.93

DOI: 10.17212/2782-2001-2021-3-53-74

Распознавание русского и индийского жестовых языков на основе машинного обучения*

М.Г. ГРИФ[1,a], Р. ЭЛАККИЯ[2,b], А.Л. ПРИХОДЬКО[1,c], М.А. БАКАЕВ[1,d], Е. РАДЖАЛАКШМИ[2,e]

[1] *630073, РФ, г. Новосибирск, пр. Карла Маркса, 20 Новосибирский государственный технический университет*
[2] *613401, Индия, Танджавур, Тамил Наду, Университет SASTRA, Школа компьютерных технологий*

[a] *grif@corp.nstu* [b] *elakkiya@cse.sastra.edu* [c] *alexeyayay@yandex.ru*
[d] *bakaev@corp.nstu.ru* [e] *rajalakshmi32210@gmail.com*

Рассматриваются подходы к распознаванию жестовых языков глухих на примере русского и индийского жестовых языков. Предлагается структура системы распознавания отдельных жестов на основе выявления пяти его компонент – конфигурации, ориентации, локализации, движения и немануальных маркеров. Приведен анализ используемых методов распознавания отдельных жестов и непрерывной жестовой речи для индийского и русского жестовых языков. Для распознавания отдельных жестов был разработан Датасет РЖЯ, включающий более 35 000 файлов. Были отобраны более тысячи жестов РЖЯ. Каждый жест исполнялся с пятью повторениями и не менее пятью глухими носителями русского жестового языка из Сибири. Для выделения эпентезы для непрерывного РЖЯ были отобраны и записаны на видео 312 предложений с пятью повторениями. Выделялись 5 типов движений: «Нет жеста», «Идет жест», «Начальное движение», «Переходное движение», «Конечное движение». Разметка предложений для выделения видов эпентезы осуществлялось на платформе Supervisely.ly. Была построена архитектура рекуррентной сети (LSTM), реализованная с помощью библиотеки машинного обучения TensorFlow Keras. Точность правильного распознавания эпентезы составила 95 %. Для индийского жестового языка были разработаны наборы данных для распознавания как отдельных жестов, так и непрерывного индийского жестового языка. Отмечается, что работа в этом направлении должна быть продолжена. Для распознавания ручных жестов применялся модуль библиотеки mediapipe holistic. Этот модуль содержит группу предварительно обученных нейросетевых алгоритмов, которые позволяют получить координаты ключевых точек тела, ладоней и лица человека на изображении. На проверочных данных удалось достичь точности 85 %. В дальнейшем необходимо значительно увеличить число размеченных данных. Для распознавания немануальных компонент был составлен ряд правил, которые со-

* *Статья получена 20 февраля 2021 г.*

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-57-45006.

ответствуют тому или иному движению частей лица. Эти правила включают состояния глаз, век, губ, языка и наклоны головы.

Ключевые слова: глухие, компьютерный сурдоперевод, русский жестовый язык, индийский жестовый язык, распознавание жестов, эпентеза, искусственная нейронная сеть, машинное обучение, наборы обучающих данных

ВВЕДЕНИЕ И ПОСТАНОВКА ЗАДАЧИ

Разработка методов компьютерного перевода национальных жестовых языков глухих в любой стране мира является важной социальной задачей, способствующей поддержке коммуникаций между глухими и слышащими. Актуальность разработки систем компьютерного перевода жестовых языков (от слышащего к глухому и наоборот) состоит как в недостаточном количестве переводчиков жестовых языков, так и в не всегда желательном посредничестве (медицина, личностные отношения и т. п.) при коммуникациях глухих и слышащих граждан [1]. В настоящее время эта задача не может считаться решенной, так как отсутствуют известные варианты машинного перевода жестовых языков, удовлетворяющие требованиям глухих. К основным требованиям глухих относятся безошибочность и качество визуализации перевода (компьютерного персонажа). Проведенные исследования среди глухих позволяют установить требования к безошибочности перевода не ниже 90 % [2] и к качеству визуализации на уровне перевода человека-переводчика. Указанные обстоятельства затрудняют коммуникации между слышащими и глухими гражданами, что является серьезной социально-экономической проблемой. Всё вышесказанное в полной мере относится к русскому (РЖЯ) и индийскому жестовым языкам. Так, например, в настоящее время при переводе русского текста на РЖЯ и наоборот используют как методы машинного обучения, так и связь грамматических систем русского звучащего языка и РЖЯ. Однако достичь необходимого уровня перевода препятствуют следующие обстоятельства:

- неполнота описания грамматической системы РЖЯ;
- отсутствие «плавности» показа жестов аватаром-переводчиком РЖЯ, система управления которого использует систему нотаций языков глухих;
- перевод осуществляется преимущественно на калькирующую жестовую речь, а не на РЖЯ, обладающий выразительными возможностями;
- высокий процент ошибок при переводе многозначных слов и омонимов на жесты РЖЯ (более 20 %);
- отсутствие значительных по объему размеченных корпусов РЖЯ (Датасетов), необходимых для реализации методов машинного перевода на основе машинного обучения;
- отсутствие надежных методов распознавания как отдельных жестов глухих, так и РЖЯ в целом.

Особую сложность представляет собой задача распознавания непрерывной жестовой речи (РЖЯ). Для успешного распознавания недостаточно только выделить отдельные жесты. Необходимо их уверенное выделение с учетом комбинаторных изменений параметров жестов, а также эпентезы.

Подробный анализ основных результатов для мировых жестовых языков содержится в [3, 4]. Так, например, в [3] был проведен анализ 396 научных

статей по тематике распознавания национальных жестовых языков с 2007 по 2017 год, каждая из которых была отнесена к категории одного из 25 жестовых языков и в дальнейшем сравнивалась по шести параметрам (методы сбора данных, статические/динамические жесты, режим записи, одноручные/двуручные жесты, методы и скорость распознавания). Анализ показал, что основные исследования по распознаванию жестовых языков были выполнены на статических, изолированных и одноручных жестах с использованием камеры.

Для распознавания [4] применялись как 2D-видеокамеры, так и браслеты и 3D-устройства (сенсор Kinect). Основные применяемые модели – нейронные сети. Несмотря на то что в отдельных случаях удавалось достичь 90 % точности распознавания, это происходило либо на ограниченном наборе жестов (50…200), либо на статичных. Что касается решения проблемы эпентезы, то применялись методы выделения признаков начала и конца жестов, включая отслеживание лица говорящего. Однако они также сегодня не могут быть признаны универсальными и надежными. Отмечается [4], что в настоящее время отсутствуют полноценные системы распознавания национальных жестовых языков.

Что касается собственно распознавания РЖЯ, то успехи здесь также невелики. Можно отметить работы [5, 6], где для распознавания РЖЯ с помощью сенсора Kinect были применены сверточные нейронные сети. Представляет также интерес разработка Датасета для 3D-моделей человека [7]. Однако нужно отметить, что Датасет был разработан для достаточно узкого диалекта РЖЯ (Санкт-Петербургского) и включал в себя небольшой набор жестов (около 300).

Для распознавания отдельных компонент жестов (конфигурация и ориентация), а также немануальных компонент был применен нейросетевой подход [8, 9]. Однако он применим в большей мере для изучения РЖЯ, а не для его распознавания. Нужно отметить также корпус РЖЯ, разработанный С.И. Бурковой [10], однако и он не может быть применен непосредственно в машинном обучении для решения задачи распознавания РЖЯ.

Таким образом, на данный момент отсутствуют достаточно полные Датасеты РЖЯ, методы распознавания используют неполный набор компонент жеста, а подходы к выделению эпентезы не являются универсальными и надежными.

Цель настоящей работы заключается в дальнейшем развитии методов распознавания русского и индийского жестовых языков на основе машинного обучения. В процессе ее достижения российские и индийские участники совместного гранта РФФИ обменивались идеями, Датасетами, разработанным программным обеспечением.

1. РАЗРАБОТКА ОБУЧАЮЩИХ ДАННЫХ ДЛЯ РУССКОГО ЖЕСТОВОГО ЯЗЫКА

Датасет русского жестового языка – это данные, которые необходимы для обучения нейронных сетей. Для анализа жестов и распознавания РЖЯ требуется много специфических данных. В настоящее время универсального Датасета для распознавания РЖЯ не существует. Так же как и для других национальных жестовых языков их недостаточно. Поэтому мы приняли ре-

шение создать собственный размеченный Датасет РЖЯ для машинного обучения. Если для разработки системы распознавания любого жестового языка применять недоработанный, неполный Датасет, то результат будет неудовлетворительным. Датасет, являющийся важнейшим элементом систем искусственного интеллекта и создаваемый с целью разработки моделей машинного обучения, требует приоритетного внимания.

При разработке Датасета выделяются три этапа: сбор жестов, разметка жестов, очистка Датасета.

В лингвистике [10] русского жестового языка используются пять уровней анализа: фонологический, морфологический, лексический, синтаксический и дискурсический. Особое внимание нужно уделять фонологическому анализу, который посвящен уровню элементарных единиц жестового языка. Как правило, в звуковых языках определена «Фонема» в связи с акустической моделью, а в жестовых языках – «Пять компонентов жеста» в связи с наличием кинематической модели руки и тела, а также с визуальной моделью мимики лица. Эта модель находится в жестовом пространстве – области, используемой говорящим для артикуляции. Вертикальная ее граница начинается чуть выше головы и заканчивается у талии, а горизонтальная граница пролегает от одного локтя до другого при свободном расположении рук.

Пять компонент жеста, элементы жестового языка, включают конфигурацию руки, ориентацию руки, локализацию, движение и немануальный компонент (рис. 1).

Рис. 1. Пять компонент жеста

Fig. 1. The five components of a gesture

Конфигурация и ориентация руки. Разные конфигурации руки являются разными формами руки при исполнении жеста. На рис. 2 даны примеры одного жеста «ножницы», где ладонь в жесте находится в разных конфигурациях, «П» и «Л» – буквы дактильного алфавита, но в одной ориентации. Разные ориентации представляют собой положения ладони в пространстве. Ладонь правой или левой руки может быть развернута вверх, вниз, вправо, влево, вверх вправо, вверх влево и в других направлениях.

Локализация. На рис. 1 локализация жеста включает два основных признака – место и сеттинг. Место исполнения – это несколько крупных областей в пределах жестового пространства: голова, лицо, шея, грудь, талия, нейтральное жестовое пространство (жест выполняется без контакта руки с

телом) и пассивная рука. Сеттинг находится внутри этой большой области. Например, место – лицо, сеттинг – правый глаз.

Движение. На рис. 1 жест «велосипед» показан стрелками как круг, который обозначает «характер» движения из признаков типа траектории (по прямой, по зигзагу, по дуге и т. п.). Второй признак – направление (вертикальный, горизонтальный и сагиттальный). Во втором типе «Локальное движение» изменяется конфигурация или ориентация руки. Например, в жесте «ножницы» (рис. 2) имеется изменение – конфигурация ладони «Л» и «П».

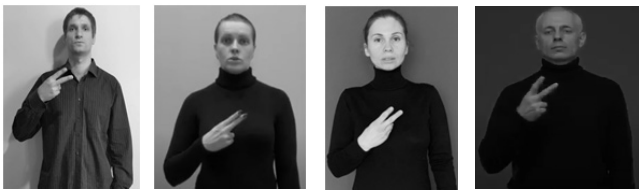


Рис. 2. Вариативность в показе жеста РЖЯ «ножницы»

Fig. 2. Variation in showing the Scissor gesture in RSL

Немануальный компонент. Немануальные компоненты включают четыре артикулятора: корпус тела, голову, плечи и мимику лица. В РЖЯ для многих жестов часто двигаются одновременно голова и корпус, изменяются мимика лица и маусинг. Под маусингом понимают движения губ говорящего на звучащем языке. Однако у говорящего на жестовом языке движения губ и языка не связаны со звучащим языком.

Типы жестов. Можно предложить следующую классификацию жестов [10]: одноручные и двуручные, статичные и динамичные, симметричные и асимметричные, синхронные и поочередные.

Формирование Датасета отдельных жестов РЖЯ. Нами был разработан Датасет РЖЯ, включающий более 35 000 жестов (изображения и видеофайлы, из них более 10 000 видеофайлов). Были отобраны более тысячи жестов РЖЯ из онлайн-словаря и книги-словаря [11, 12]. Каждый жест исполнялся с пятью повторениями и не менее пятью глухими носителями русского жестового языка из Сибири.

Сбор и запись Датасета осуществлялась при выполнении следующих условий (требований):

- на чистом фоне говорящий в жестовом пространстве был одет в черный свитер;
- жесты показывали только носители русского жестового языка (говорящие на РЖЯ);
- не менее пяти повторений жеста каждым носителем РЖЯ;
- не менее 20 повторений каждого жеста различными говорящими на РЖЯ;
- до включения видеозаписи говорящий должен молча стоять в жестовом пространстве, после включения видеозаписи говорящий молчит 2-3 секунды, потом говорит, затем молчит 2-3 секунды до выключения;
- не менее 1000 воспроизводимых видеожестов;
- разрешение изображения не менее 1920×1080, частота не менее 30 кадров в секунду.

Такие действия должны обеспечить полноту и информативность в формируемом Датасете. Однако на практике мы столкнулись с наличием у говорящих мелких фонологических привычек показа жестов, что приводило к несовпадениям со словарем РЖЯ (например, разные конфигурации и ориентации руки, а также сеттинг в области локализации жеста).

Качество разметки данных зависит и от качества работы оператора по разметке жестов для машинного обучения. Разметка жестов проводилась на платформе Supervisely.ly [13]. Выделялась область ладоней и элементов лица. После сбора жестов для удобства хранения применялась система имен видеофайлов, включающая номера жеста, человека и повторения (например, имя файла 43_3_3.mp4 – номер жеста 43, 3-й номер человека в списке информантов и 3-е повторение).

Формирование Датасета для выделения эпентезы в РЖЯ. В [4] приведен обзор работ, связанных с выделением эпентезы в мировых жестовых языках. Под эпентезой понимается движение между соседними жестами [10]. Для успешного распознавания РЖЯ необходимо отделить эпентезу от собственно жестов. Основная идея заключается в выделении признаков начала следующего жеста или паузы в речи. Это может быть состояние покоя говорящего, положения рук, взгляда и т. п. [14]. Применяют также приемы вычленения быстрых и медленных движений в видеопотоке. Если процесс жестовой речи свести к анализу последовательности выявленных компонент (конфигурация – ориентация ладони, локализация, траектории рук и характер движения, немануальные компоненты), то каждое такое событие станет «кандидатом» на начало распознавания следующего жеста. При этом может оказаться, что процесс нового поиска уже начат, хотя не окончен предыдущий. Подобный подход к распознаванию жестовой речи является достаточно трудоемким. Представляется, что его можно модернизировать с введением функции принадлежности (или некоторой вероятности) каждого события к началу нового жеста. В этом случае алгоритм распознавания нового жеста можно запускать только после некоторых, наиболее вероятных событий. Другой подход состоит в выделении всех элементов жестовой речи: собственно жестов, состояния покоя говорящего, комбинаторных изменений жестов и эпентезы. В этом случае необходимо расширить Датасет для распознавания жестов и другими элементами, связанными с комбинаторными изменениями жестов. На данный момент сложно оценить необходимый объем дополнений. С учетом большого числа типов комбинаторных изменений жестов он может оказаться значительным.

Для выделения эпентезы в РЖЯ мы использовали следующие понятия: «начальное движение», «конечное движение» и «промежуточное движение». «начальное» – это подготовительное движение, «конечное» – когда рука возвращается в исходное положение. Считаем, что здесь имеется аналогия терминов из лингвистики звуковых языков, обозначающих фазы артикуляции звука. Здесь, по сути, то же самое:

- экскурсия (подготовка органов к артикуляции);
- выдержка (фаза артикуляции);
- рекурсия (возврат органов артикуляции в исходное положение).

Были отобраны и записаны на видео 312 предложений с пятью повторениями на РЖЯ от двух глухих носителей РЖЯ (160 предложений у женщины и 152 – у мужчины), из которых 142 – уникальные предложения (рис. 3).

129	ваш (04)	дети (81)	упрямый_2	не.выносить	артикуляция_1
130	осень	береза	лист	изменять	оранжевый_2
131	представтель_1	сообщение	премия	прекращать	причина
132	мышь	серый	слон	испуганный	дрожать_1
133	суп	недостаток	соль_1 (58)	перец (45)	
134	нравится	деревня (80)	дышать (14)	воздух	чистый
135	ванная	балкон	нет		
136	зима	февраль	дрожать_2	валенки (114)	теплый
137	желание	выполнить_2	лето	самолет	отпуск
138	смотреть.телевизор	показывать	история	источник	земля (91)
139	литература	книга	запрещать	копировать_2	будет
140	военный_1	воля (06)	воевать	до_1	победа
141	показывать	фильм (68)	тайна	фашизм	страх
142	мой	бизнес (102)	сам	идея	
143					

Рис. 3. Примеры уникальных предложений на РЖЯ

Fig. 3. Examples of unique sentences in RSL

В каждом предложении выделены жесты РЖЯ, описанные ранее (рис. 4).

Рис. 4. Запись предложений на РЖЯ

Fig. 4. Recording sentences using RSL

Разметка предложений для выделения видов эпентезы (начальное, конечное и промежуточное движение) осуществлялась на платформе Supervisely.ly [13] (рис. 5).

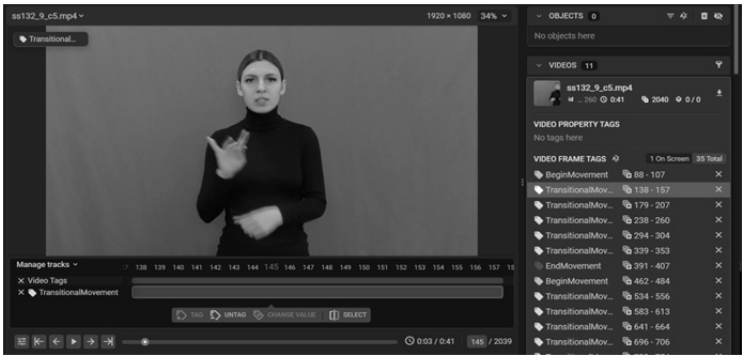


Рис. 5. Инструмент по разметке на платформе Supervise.ly

Fig. 5. Data labeling tool powered by Supervise.ly

Были введены метки для выделенных движений. На рис. 5 отмечены кадры в видеофайле с конкретным типом движения. Результат разметки помещается в json-файлы. Так, например, на рис. 5 справа показаны три типа меток: начальное движение [88:107], переходное [138:157], [179:207],

[238:260], [294:304] и конечное [391:407]. Для удобства были расширены варианты движений: «нет жеста», «идет жест», «начальное движение», «переходное движение», «конечное движение». В результате имеем 5 полных выходных классов на видеофайлах с жестами.

Для обучения нейронной сети нужна дополнительная разметка выделенных движений скелетной моделью. Для этого была использована система MediaPipe Pose [15]. MediaPipe Pose – это решение машинного обучения для точного отслеживания позы тела, позволяющее вывести 33 трехмерных ориентира на всем теле из видеокадров RGB с использованием инструмента BlazePose. После извлечения этих данных создаются numpy-файлы, включающие информацию о координатах ключевых точек тела. На входе подается видео, а на выходе имеем numpy-файл, содержащий в себе массив нормализованных координат для каждой ключевой точки руки. Размерность выходного массива равна $F \times C$, где F – количество фреймов в видео, C – количество анализируемых точек, 33×2 (рис. 6).

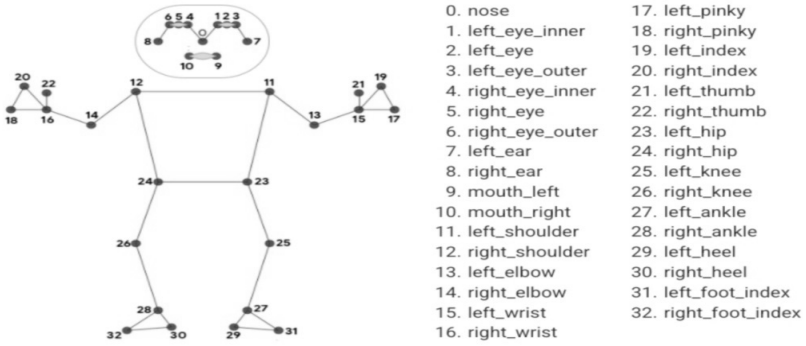


Рис. 6. Ключевые точки скелетной модели

Fig. 6. Key points of the skeletal model

Практическая реализация многослойных персептронов необходимой архитектуры и конфигурации производилась с помощью библиотеки машинного обучения TensorFlow Keras на языке программирования Python 3.6. Сеть, изображенная на рис. 7, представляет собой однослойный персептрон с 512 нейронами, всего имеется 1 120 773 параметров, которые не оптимизируются в процессе обучения. Максимальное количество эпох обучающего цикла составляет 100.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_10 (LSTM)               (None, 512)               1118208
_________________________________________________________________
dense_10 (Dense)             (None, 5)                 2565
=================================================================
Total params: 1,120,773
Trainable params: 1,120,773
Non-trainable params: 0
```

Рис. 7. Архитектура рекуррентной сети (LSTM), реализованная с помощью библиотеки машинного обучения TensorFlow Keras

Fig. 7. The recurrent network architecture (LSTM) implemented using the TensorFlow Keras machine learning library

Результаты распознавания типов движений для выделения эпентезы представлены на рис. 8. Точность правильного распознавания составила 95 %.

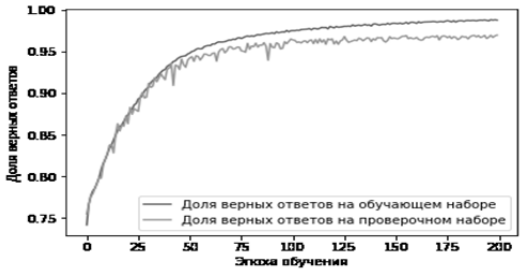
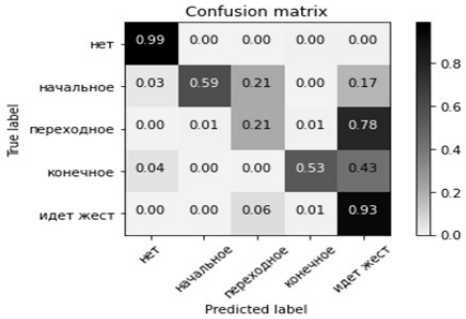


График процесса обучения



Матрица отклонений

Рис. 8. Результаты распознавания типов движений для выделения эпентезы

Fig. 8. The results of recognition of movement types to highlight epenthesis

2. РАЗРАБОТКА ОБУЧАЮЩИХ ДАННЫХ И КОРПУСОВ ДЛЯ ИНДИЙСКОГО ЖЕСТОВОГО ЯЗЫКА

В настоящее время отсутствуют общедоступные наборы данных по индийским жестам для разработки системы распознавания индийского жестового языка (ISLR). Поэтому мы собрали и создали наборы данных для распознавания как изолированного, так и непрерывного индийского жестового языка (ISL). Для изолированного ISL были собраны и созданы два набора данных, а именно ISLAN и ISLW. Так, ISLAN состоит из ISL-представлений английского алфавита и чисел; ISLW состоит из ISL-представлений 2774 английских слов. Для непрерывного ISL был создан набор данных, а именно ISLS, который состоит из ISL-представлений ста простых английских предложений. При создании этих наборов данных использовались жесты как для одной руки, так и для двух рук.

Алфавит. ISLAN – это коллекция из 700 изображений и 24 видеороликов. Изображения и видео из набора данных ISLAN были подготовлены с помощью ученых-исследователей и студентов Университета SASTRA Deemed, Тамил Наду (Индия). Цветные изображения с необработанными метками (формат JPG) и видео (формат MP4) были записаны со смартфона. Были задействованы шесть носителей языка (трое мужчин и три женщины) с различным оттенком кожи и с разными размерами рук [16]. Были использованы как одноручные, так и двуручные жесты.

На рис. 9 показано двуручное обозначение буквы A, представленное шестью разными говорящими.

Рис. 9. Изображение двуручного жеста говорящего для буквы A

Fig. 9. Images of the two-handed speaker for the finger spelling of the letter A

На рис. 10 показано одноручное обозначение буквы A, представленное шестью разными говорящими. На рис. 11 показано обозначение цифровой буквы A, представленное пятью разными говорящими.

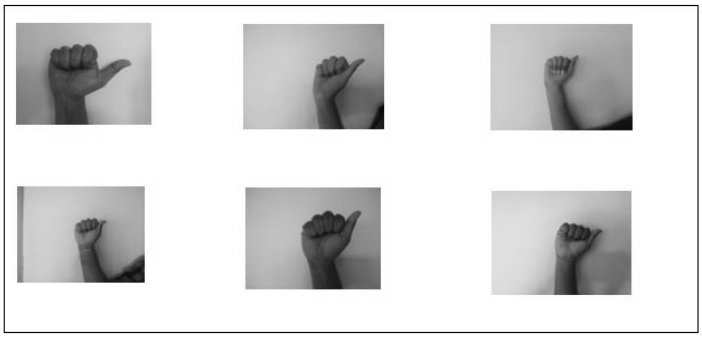


Рис. 10. Изображение жестов, представляющих одноручное лицо с несколькими говорящими сторонами для буквы A

Fig. 10. Gestures representing a one-handed person with several speakers for the finger spelling of the letter A

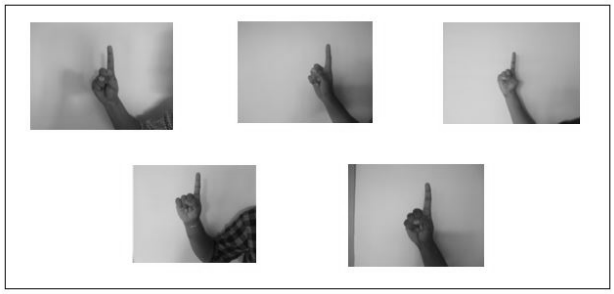


Рис. 11. Изображения представления нескольких говорящих лиц для числа 1

Fig. 11. Sign images of several speakers for the number 1

Отдельные жесты. В основу набора данных ISLW (отдельные жесты) был положен общедоступный портал индийского жестового языка [17].

ISLW состоит из коллекции 2774 индийских видео с несколькими говорящими. Каждый жест в ISLW представляет собой видео в алфавитном порядке. Видео имеет разрешение изображения 480 пикселей с частотой 30 кадров в секунду. Все видео в формате MP4, а кадры изображений – в формате JPEG. Видео были сняты с неоднородным фоном и с обычными условиями освещения в формате RGB в закрытом помещении. На рис. 12 показано жестовое представление слова «Аббревиатура» в индийском жестовом языке.

Рис. 12. Индийское обозначение жеста «Аббревиатура»

Fig. 12. An Indian sign designation of the Abbreviation gesture

Предложения. Мы также создали коллекцию индийских жестовых непрерывных предложений, а именно ISLS, который представляет собой полностью маркированный большой словарный набор видеоданных на уровне предложений на индийском жестовом языке. Каждое видео показывает перевод на язык жестов простого предложения на английском языке. Был создан набор данных на основе представлений говорящих лиц как одной, так и двумя руками. Все видео в формате MP4, а кадры изображений – в формате JPEG.

Набор видеоданных индийского непрерывного распознавания жестового языка для предложений (ISLS) был создан с помощью ученых-исследователей и студентов Университета SASTRA Deemed, Тамил Наду (Индия). Видео были сняты с помощью цифровой зеркальной камеры, прикрепленной к штативу, на неоднородном фоне и при хорошем освещении. Видео были сняты в формате RGB и в помещении с частотой кадров 25 кадров в секунду и с разрешением 1920 × 1080. Набор данных ISLS состоит из жестовых представлений ста уникальных простых предложений, выполненных семью говорящими лицами, среди которых две женщины, а остальные – мужчины. Для сбора этого набора данных были привлечены люди с разным оттенком кожи, а также с разным размером рук. ISLS станет первым набором видеоданных на индийском жестовом языке, состоящим из 100 простых предложений на английском языке. ISLS состоит из 700 видеороликов и 35 299 кадров изображений. Из этих 700 жестов были извлечены видеопредложения, состоящие из 186 фотожестов. Общее количество кадров на один фотожест варьируется. Кадры, извлеченные из видео, имели разные вариации, а также размеры, поэтому они были предварительно обработаны для создания кадров изображений одинакового размера 64 × 64. На рис. 13 показано жестовое представление предложения «Сколько вам лет?».

Рис. 13. Индийский жест, обозначающий предложение «Сколько Вам лет?»

*Fig. 13.*An Indian gesture showing a sentence “How old are you?”

Пути развития. Сбор и создание этих наборов данных окажется полезным для исследователей, работающих над ISLR на основе видеокамеры. ISLAN, ISLW и ISLS будут очень полезны исследователям для разработки новых методов и повышения производительности уже существующих систем ISLR. Эти наборы данных также повысят осведомленность людей об изучении языка жестов. Набор данных можно расширить, пытаясь сделать фон однородным и с различной интенсивностью света. Кроме того, мы также можем попытаться захватить изображения под разными углами или воспроизвести и выполнить различные методы преобразования изображений, чтобы расширить набор данных для дальнейшего обучения. Указанные наборы данных мы планируем расширить, так как их недостаточно для покрытия большого словарного запаса, имеющегося в разговорных языках. В ISLR всё еще есть много проблем, но основным недостатком является отсутствие общедоступных наборов данных. Несмотря на то что доступно несколько наборов данных, в них всё еще не учтены многие ограничения, такие как несколько говорящих лиц, большой словарный запас, интенсивность света изображения, расстояние от камеры, угол камеры, жесты, выполняемые несколькими руками и одной рукой, оттенок кожи, фон и т. д. На данный момент мы рассмотрели только три основные функции: несколько говорящих лиц, оттенок кожи, жесты одной и двумя руками. Развитие методов глубокого обучения в области компьютерного зрения позволило исследователям создавать приложения для распознавания языка жестов. Собранные и созданные наборы данных помогут предоставить ресурсы в форме общедоступного набора данных ISL для дальнейших исследований в ISLR. Мы также благодарим Департамент науки и технологий (DST) Индии за финансовую поддержку в рамках Российско-индийского совместного проекта (INT/RUS /RFBR/393). Мы также благодарим SASTRA Deemed University (Танджавур, Индия) за расширение инфраструктурной поддержки для проведения этой исследовательской работы.

3. РАСПОЗНАВАНИЕ ОТДЕЛЬНЫХ ЖЕСТОВ РУССКОГО ЖЕСТОВОГО ЯЗЫКА

В основу алгоритмов взят подход к распознаванию отдельных жестов по наличию определенного набора компонент: конфигураций-ориентаций, локализации, траектории и характера движения, а также немануальных компонент.

Распознавание ручных жестов

Одним из методов решения задачи распознавания жестов был взят модуль библиотеки mediapipe holistic [18]. Этот модуль содержит группу предварительно обученных нейросетевых алгоритмов, которые позволяют получить координаты ключевых точек тела, ладоней и лица человека на изображении.

Использование модуля mediapipe holistic позволяет решить проблемы с получением координат точек ладоней в различных условиях записи видео (разная освещенность, разрешение и качество видеозаписи и пр.). Для дальнейшей работы с классификацией жеста используется часть результата работы модуля mediapipe – координаты левой и правой ладони (21 точка для каждой руки, координата каждой точки имеет вид $\{x_n, y_n, z_n\}$) человека в кадре. Координаты лица и тела в дальнейшем применяются для учета в модели неманульного компонента РЖЯ. Для дальнейшей классификации используются именно координаты точек каждой из ладоней.

На текущем этапе разработки имеется ограниченное число размеченных данных, при этом имеется большое число цифровых изображений и видеозаписей, на которых показаны те или иные жесты РЖЯ.

Было принято решение использовать эти неразмеченные данные для повышения качества работы модели классификации жестов.

Первым этапом обучения модели классификатора является обучение разреженного автокодировщика (sparse autoencoder) [19], цель которого на основании неразмеченных данных обучиться преобразовывать представление ладоней в виде координат в более сложное представление для классификации семантически сложных признаковых пространств.

Координаты ладоней представляют собой 21 трехмерный вектор $[\{x_1, y_1, z_1\}\ldots\{x_n, y_n, z_n\}]$, который может быть преобразован в 63 признака для входа нейронной сети. Стоит отметить, что помимо координат точек ладоней можно выделить большее число признаков (например, связь точек ладони). Так, точка кончика пальца соединена с точкой конца средней фаланги, а та, в свою очередь, с точкой конца проксимальной фаланги пальца. Еще одним примером может служить тот факт, что точка основания пальца всегда будет на относительно большом (относительно всей ладони) расстоянии в трехмерном пространстве. Еще большее число подобных примеров можно вывести исходя из эмпирических наблюдений за ладонями людей, и в частности за положениями ладоней в РЖЯ (рис. 14).

Гипотетически такие дополнительные признаки можно генерировать на основании большого числа представлений, количество которых около одного миллиона среди имеющихся в нашем распоряжении данных. За счет регуляризации скрытого представления (latent space) в разреженном автокодировщике (sparse autoencoder) [19] появляется возможность генерировать необходимые дополнительные признаки.

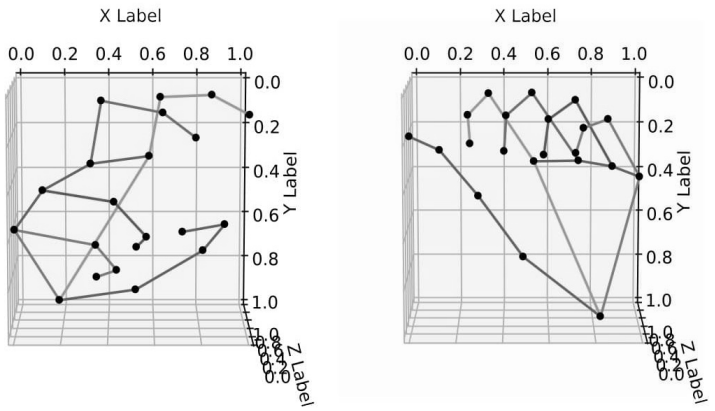


Рис. 14. Пример точек ладоней

Fig. 14. An example of palm points

Таким образом, первым этапом обучения модели является обучение автокодировщика на корпусе неразмеченных данных. Обучение автокодировщика позволило достичь точности 97,5 %. Модель кодировщика имеет следующую структуру.

Слой (type)	Output Shape	Param #
========== Слои кодировщика ==========		
input_1 (InputLayer)	[(None, 3, 21)]	0
flatten (Flatten)	(None, 63)	0
dense (Dense)	(None, 256)	16384
dense_1 (Dense)	(None, 512)	131584
batch_normalization (BatchNo	(None, 512)	2048
dense_2 (Dense)	(None, 1024)	525312
========== Слои декодировщика ==========		
input_2 (InputLayer)	(None, 1024)	
dense_3 (Dense)	(None, 1024)	
batch_normalization (BatchNo	(None, 1024)	
dense_4 (Dense)	(None, 512)	
dense_5 (Dense)	(None, 256)	
dense_6 (Dense)	(None, 63)	
reshape (Reshape)	[(None, 3, 21)]	

В результате обучения такого кодировщика число признаков было увеличено до 1024 в скрытом представлении. Для дальнейшего обучения классификатора используется только первая часть автокодировщика (autoencoder) – кодировщик (encoder). При дальнейшем обучении эта часть нейронной сети остается с неизменными значениями весов.

Следующим этапом обучения модели классификации является обучение на размеченных данных. Обучение выполняется на нескольких наборах данных, собранных нами самостоятельно.

Введен параметр минимального числа примеров для обучения, принятый равным 40. В случае если минимальное число примеров класса не достигнуто, метка класса становится '0' (не относится к распознаваемым жестам). В настоящее время число распознаваемых жестов равно 268.

Используемая базовая модель классификатора имеет следующий вид.

Слой (type)	Output Shape	Param #
========== Слои кодировщика ==========		
input_1 (InputLayer)	[(None, 3, 21)]	0
flatten (Flatten)	(None, 63)	0
dense (Dense)	(None, 256)	16384
dense_1 (Dense)	(None, 512)	131584
batch_normalization (BatchNo	(None, 512)	2048
dense_2 (Dense)	(None, 1024)	525312
========== Слои классификации ==========		
dense (Dense)	(None, 1024)	1049600
dense_1 (Dense)	(None, 2048)	2099200
dense_2 (Dense)	(None, 1024)	2098176
batch_normalization (BatchNo	(None, 1024)	4096
dense_3 (Dense)	(None, 53)	54325

На проверочных данных удалось достичь точности в 81 %.

Классификатор жеста. Использовалась обучающая выборка из 2195 видеозаписей, содержащая 774 отдельных жеста, имеющих подпись, состоящую из последовательностей трех жестов для левой и правой руки. Максимальная длина видеозаписи в наборе данных – 184 кадра. Для обучения модели классификации жестов применялась ранее обученная модель для распознавания конфигурации-ориентации жеста (МКОЖ) на изображениях 268 возможных жестов (по 134 для каждой руки). В дальнейшем коли-

чество поддерживаемых жестов будет увеличиваться с ростом размера набора данных.

Предварительная обработка видеозаписей. Выполняется обработка каждого кадра ранее обученной моделью для определения конфигурации-ориентации каждой ладони в каждом кадре видеозаписи. Таким образом, видео преобразуется в две последовательности данных вида

Xlh = ['0', '0', '0', '0', …, 's15d31', 's15d31', 's18d00', 's15d31', 's14c51', '0', 's1dc30', 's1dc30', 's1dc30', 's15d32', '0', ..., '0', 's1dc30', 's15d32', '0', 's13f30', 's18c08', '0', '0', '0'] (для левой и правой руки отдельно).

Здесь '0' – отсутствие руки в кадре / модель распознавания конфигурации ориентации не обучена на этом жесте / модель не распознала жест; 's15d31' – распознанная моделью МКОЖ конфигурация-ориентация. Общая длина такой последовательности соответствует количеству кадров в видеозаписи жеста. На текущий момент принята максимальная длина видеозаписей – 200 кадров, что соответствует 6,6 секунды при частоте кадров видеозаписи, равной 30. Модель классификации видео (МКВ) в дальнейшем работает с данными вида Xlh. Если количество элементов в полученной последовательности менее 200, то выполняется ее «растяжение», новые элементы последовательности подбираются одним из следующих возможных способов:

1) заполнение левого и правого края «0»;

2) заполнение правого края «0»;

3) интерполяция, заполнение копиями рядом стоящих элементов, т. е. [«0», «s12000», «s20100»] → [«0», «0», «s12000», «s12000», «s20100», «s20100»].

Наиболее эффективным в ходе обучения модели был способ 3.

Далее выполняется преобразование последовательностей к матричному виду:

```
[ [0.          0.        0.        0.        0.      ]
  [0.          0.        0.        0.        0.      ]
  [0.          0.        0.        0.        0.      ]
  ...
  [0.5011157   0.28710333  0.8312101   0.11121226  0.      ]
  [0.4776144   0.30229354  0.6563805   0.04033527  0.      ]
  ...
  [0.          0.        0.        0.        0.      ]].
```

В этом виде конфигурация-ориентация преобразуется из вида «s12000» в вектор $[x_1, x_2, x_3, x_4, x_5]$, где $x_i = \dfrac{x_i}{x_{i,\max} - x_{i,\min}}$.

Такое представление обусловлено введенной структурой [20], где жест может быть интерпретирован как вектор размерности 5 (по одному измерению на каждый элемент формальной записи конфигурации-ориентации)

Таким образом, видео преобразуется в последовательность векторов, описывающих ориентацию-конфигурацию, на которых выполняется обучение нейронной сети, архитектура сети показана на рис. 15.

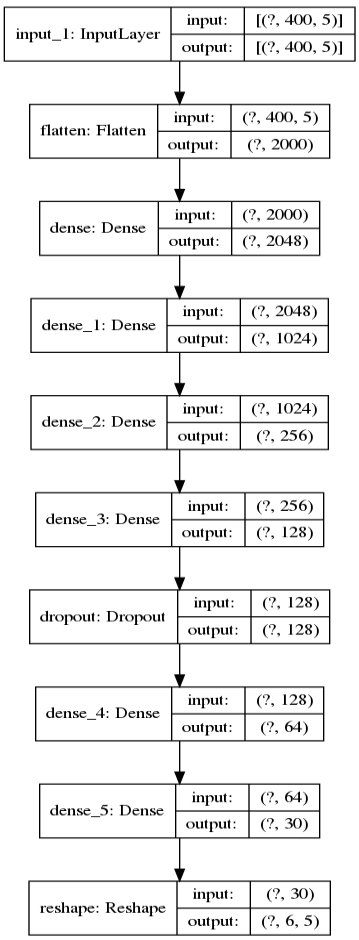


Рис. 15. Архитектура нейронной сети

Fig. 15. Neural network architecture

Результатом работы нейронной сети является матрица вида

```
[  [0.          0.        0.        0.        0.        ]
   [0.          0.        0.        0.        0.        ]
[0.          0.        0.        0.        0.        ]
[0.5         0.26666667  0.8       0.        0.13333333]
[0.5         0.26666667  0.8       1.        0.        ]
[0.          0.        0.        0.        0.        ]].
```

То есть по три конфигурации-ориентации на каждую руку.

Далее в качестве меры близости полученной последовательности векторов используется евклидово расстояние. Результат работы модели показывает точность около 80 %, однако является зависимым от качества модели определения конфигурации-ориентации ладони. С ростом качества модели МКОЖ улучшится и качество работы данной модели.

Распознавание немануальных компонент. Использование модуля mediapipe holistic позволяет решить проблемы с получением координат точек

лица в различных условиях записи видео (разная освещенность, разрешение и качество видеозаписи и пр.). Для работы с классификацией немануального компонента используется часть результата работы модуля mediapipe – координаты точек лица (468 точек лица, координаты каждой из точек имеют вид $\{x_n, y_n, z_n\}$) человека в кадре.

В качестве интересующих точек для контроля движений губ и глаз выбраны точки, указанные на изображении (рис. 16).

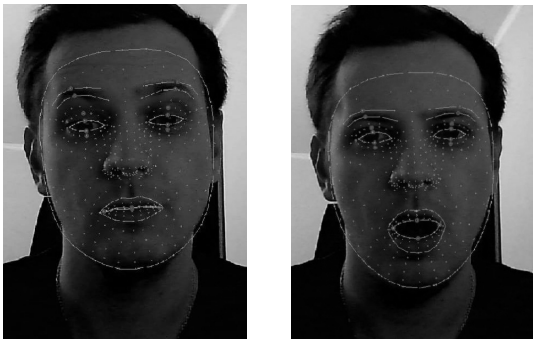


Рис. 16. Распознавание немануальных компонент

Fig. 16. Recognition of non-manual components

На каждом кадре выполняется вычисление эвклидова расстояния между точками лица, и эмпирическим путем выявляется факт движения. Например, поднятие бровей – это увеличение расстояния между нижним краем брови и верхним веком. Таким образом, составлен ряд правил, которые соответствуют тому или иному движению частей лица. Эти правила включают состояния глаз, век, рта, языка и наклоны головы.

ЗАКЛЮЧЕНИЕ

Были рассмотрены подходы к распознаванию жестовых языков глухих на примере русского и индийского жестовых языков. Проведен анализ используемых методов распознавания отдельных жестов и непрерывной жестовой речи для индийского и русского жестовых языков. Для распознавания отдельных жестов был разработан Датасет РЖЯ, включающий более 35 000 файлов. Для выделения эпентезы в непрерывном РЖЯ были отобраны и записаны на видео 312 предложений с пятью повторениями. Выделено пять типов движений: «нет жеста», «идет жест», «начальное движение», «переходное движение», «конечное движение». Разметка предложений для выделения видов эпентезы осуществлялась на платформе Supervisely.ly. Была построена архитектура рекуррентной сети (LSTM), реализованная с помощью библиотеки машинного обучения TensorFlow Keras. Точность правильного распознавания эпентезы составила 95 %. Для индийского жестового языка были разработаны наборы данных по распознаванию как отдельных жестов, так и непрерывного жестового языка. Для распознавания ручных жестов применялся модуль библиотеки mediapipe holistic. Этот модуль содержит группу предварительно обученных нейросетевых алгоритмов, которые поз-

воляют получить координаты ключевых точек тела, ладоней и лица человека на изображении. На проверочных данных удалось достичь точности в 85 %. Для распознавания немануальных компонет был составлен ряд правил, которые соответствуют тому или иному движению частей лица. Эти правила включают состояния глаз, век, рта, языка и наклоны головы.

СПИСОК ЛИТЕРАТУРЫ

1. Лексические и грамматические аспекты разработки компьютерного сурдопереводчика русского языка / М.Г. Гриф, О.О. Королькова, Л.Г. Панин, М.К. Тимофеева, Е.Б. Цой. – Новосибирск: Изд-во НГТУ, 2013. – 292 с.

2. *Гриф М.Г. Королькова О.О., Мануева Ю.С.* Машинный перевод русского жестового языка глухих // Информатика: проблемы, методы, технологии: материалы 20 международной научно-методической конференции, Воронеж, 13–14 февр. 2020. – Воронеж, 2020. – С. 1591–1597. – ISBN 978-5-6042216-9-3.

3. *Wadhawan A., Kumar P.* Sign language recognition systems: a decade systematic literature review // Archives of Computational Methods in Engineering. – 2021. – Vol. 28. – P. 785–813. – DOI: 10.1007/s11831-019-09384-2.

4. Распознавание русского и индийского языков жестов глухих / Р. Элаккия, М.Г. Гриф, А.Л. Приходько, М.А. Бакаев // Научный вестник НГТУ. – 2020. – № 2–3 (79). – С. 57–76. – DOI: 10.17212/1814-1196-2020-2-3-57-76. – Текст англ.

5. Sign language numeral gestures recognition using convolutional neural network / I. Gruber, D. Ryumin, M. Hrúz, A. Karpov // Interactive Collaborative Robotics. – Cham: Springer, 2018. – P. 70–77.

6. *Константинов В.М., Орлова Ю.А., Розалиев В.Л.* Разработка 3D-модели тела человека с использованием MS Kinect // Известия Волгоградского государственного технического университета. – 2015. – № 6 (163). – С. 65–69.

7. TheRuSLan: Database of Russian Sign Language / I. Kagirov, D. Ivanko, D. Ryumin, A. Axyonov, A. Karpov // Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020). – Marseille, 2020. – P. 6079–6085.

8. Automatic classification of handshapes in Russian sign language / M. Mukushev, A. Imashev, V. Kimmelman, A. Sandygulova // Proceedings of the LREC 2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives. – Marseille, 2020. – P. 165–169.

9. Eyebrow position in grammatical and emotional expressions in Kazakh-Russian Sign Language: A quantitative study / V. Kimmelman, A. Imashev, M. Mukushev, A. Sandygulova // Plos One. – 2020. – Vol. 15 (6). – DOI: 10.1371/journal.pone.0233731.

10. Введение в лингвистику жестовых языков. Русский жестовый язык: учебник / ред.: С.И. Буркова, В.И. Киммельман. – Новосибирск: Изд-во НГТУ, 2019. – 356 с.

11. Словарь русского жестового языка. – URL: https://www.spreadthesign.com/ru (дата обращения: 31.08.2021).

12. *Базоев В.З.* Словарь русского жестового языка. – М.: Флинта, 2009. – 528 с.

13. Supervisely – веб-платформа для компьютерного зрения. – URL: https://supervise.ly/ (дата обращения: 31.08.2021).

14. *Кебец П.Л.* Маркеры перехода хода в дискурсе русского жестового языка // Материалы Седьмой конференции по типологии и грамматике для молодых исследователей (г. Санкт-Петербург, 4–6 ноября 2010 г.). – СПб.: Наука, 2010. – С. 75–80.

15. Система MediaPipe Pose. – URL: https://google.github.io/mediapipe/solutions/pose.html (дата обращения: 31.08.2021).

16. *Elakkiya R., Rajalakshmi E.* ISLAN // Mendeley Data. – 2021. – Ver. 1. – DOI: 10.17632/rc349j45m5.1.

17. Словарь индийского языка жестов. – URL: http://indiansignlanguage.org/dictionary/ (дата обращения: 31.08.2021).

18. Mediapipe: a framework for building perception pipelines / C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M.G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, M. Grundmann. – arXiv preprint arXiv:1906.08172. – 2019.

19. *Le Q.V.* Building high-level features using large scale unsupervised learning // 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. – Vancouver, BC, Canada, 2013. – P. 8595–8598.

20. Formal SignWriting (draft-slevinski-formal-signwriting-07). – URL: https://tools.ietf.org/id/draft-slevinski-formal-signwriting-07.html (accessed: 31.08.2021).

Гриф Михаил Геннадьевич, профессор кафедры автоматизированных систем управления факультета автоматики и вычислительной техники Новосибирского государственного технического университета. Сфера научных интересов – проектирование сложных систем и систем компьютерного перевода на язык жестов. Опубликовано более 300 научных и учебных работ. E-mail: grif@corp.nstu.nstu.ru

Элаккия Р., AP-III, Школа компьютерных технологий, CSE, Университет SASTRA Deemed, Индия. Научные интересы – методы распознавания жестовых языков глухих. Опубликовано более 30 научных и учебных работ. E-mail: elakkiya@cse.sastra.edu

Приходько Алексей Леонидович, младший научный сотрудник кафедры автоматизированных систем управления факультета автоматики и вычислительной техники Новосибирского государственного технического университета. Научные интересы – методы распознавания жестовых языков глухих. Опубликовано 15 научных работ. E-mail: alexeyayay@yandex.ru

Бакаев Максим Александрович, доцент кафедры автоматизированных систем управления факультета автоматики и вычислительной техники Новосибирского государственного технического университета. Научные интересы: взаимодействие человека с компьютером, дизайн интерфейса, машинное обучение. Имеет более 100 публикаций. E-mail: bakaev@corp.nstu.ru

Раджалакшми Е., аспирант, Школа компьютерных технологий, CSE, Университет SASTRA Deemed, Индия. Область научных интересов: рекомендательные системы, умные города, машинное обучение. Опубликовано 3 научные и учебные работы.

Grif Mikhail G., professor at the Department of Automated Control Systems, Faculty of Automation and Computer Engineering, Novosibirsk State Technical University. His research interests include design of complex systems and computer sign language translation systems. He has published more than 300 scientific and tutorials. E-mail: grif@corp.nstu.nstu.ru

Elakkiya R., AP-III, School of Computing, CSE, SASTRA Deemed University, India. His research interests: cover methods for recognizing sign languages of the deaf. He has published more than 30 scientific and tutorials. E-mail: elakkiya@cse.sastra.edu

Prikhodko Alexey L., junior researcher at the Department of Automated Control Systems, Faculty of Automation and Computer Engineering, Novosibirsk State Technical University. His research interests include methods for recognizing sign languages of the deaf. He has published 15 research papers. E-mail: alexeyayay@yandex.ru

Bakaev Maxim A., associate professor at the Department of Automated Control Systems, Faculty of Automation and Computer Engineering, Novosibirsk State Technical University. His research interests are human-computer interaction, interface design, and machine learning. He has over 100 publications. E-mail: bakaev@corp.nstu.ru

Rajalakshmi E., graduate student, School of Computing, CSE, SASTRA Deemed University, India. His research interests cover: recommender systems, smart cities, and machine learning. He has published 3 research papers and tutorials. E-mail: rajalakshmi32210@gmail.com

DOI: 10.17212/2782-2001-2021-3-53-74

Recognition of Russian and Indian sign languages based on machine learning[*]

M.G. GRIF[1,a], *R. ELAKKYA*[2,b], *A.L. PRIKHODKO*[1,c], *M.A. BAKAEV*[1,d], *E. RAJALAKSHMI*[2,e]

[1] *Novosibirsk State Technical University, 20, K. Marx Prospekt, Novosibirsk, 630073, Russian Federation*
[2] *613401, India, Thanjavur, Tamil Nadu, SASTRA Deemed University, School of Computing, CSE*

[a] *grif@corp.nstu* [b] *elakkiya@cse.sastra.edu* [c] *alexeyayay@yandex.ru*
[d] *bakaev@corp.nstu.ru* [e] *rajalakshmi32210@gmail.com*

Abstract

In the paper, we consider recognition of sign languages (SL) with a particular focus on Russian and Indian SLs. The proposed recognition system includes five components: configuration, orientation, localization, movement and non-manual markers. The analysis uses methods of recognition of individual gestures and continuous sign speech for Indian and Russian sign languages (RSL). To recognize individual gestures, the RSL Dataset was developed, which includes more than 35,000 files for over 1000 signs. Each sign was performed with 5 repetitions and at least by 5 deaf native speakers of the Russian Sign Language from Siberia. To isolate epenthesis for continuous RSL, 312 sentences with 5 repetitions were selected and recorded on video. Five types of movements were distinguished, namely, "No gesture", "There is a gesture", "Initial movement", "Transitional movement", "Final movement". The markup of sentences for highlighting epenthesis was carried out on the Supervisely.ly platform. A recurrent network architecture (LSTM) was built, implemented using the TensorFlow Keras machine learning library. The accuracy of correct recognition of epenthesis was 95 %. The work on a similar dataset for the recognition of both individual gestures and continuous Indian sign language (ISL) is continuing. To recognize hand gestures, the mediapipe holistic library module was used. It contains a group of trained neural network algorithms that allow obtaining the coordinates of the key points of the body, palms and face of a person in the image. The accuracy of 85 % was achieved for the verification data. In the future, it is necessary to significantly increase the amount of labeled data. To recognize non-manual components, a number of rules have been developed for certain movements in the face. These rules include positions for the eyes, eyelids, mouth, tongue, and head tilt.

Keywords: Deaf, computer sign language translation, Russian sign language, Indian sign language, gesture recognition, epenthesis, artificial neural network, machine learning, training datasets

REFERENCES

1. Grif M.G., Korolkova O.O., Panin L.G., Timofeeva M.K., Tsoy E.B. *Leksicheskie i grammaticheskie aspekty razrabotki komp'yuternogo surdoperevodchika russkogo yazyka* [Lexical and grammatical aspects of developing computer Russian sign language translator]. Novosibirsk, NSTU Publ., 2013. 292 p.

2. Grif M.G. Korolkova O.O., Manueva Yu.S. [Machine translation of Russian sign language for the deaf]. *Informatika: problemy, metody, tekhnologii* [Informatics: problems, methods, technologies]. Materials of the 20th International scientific conference, Voronezh, 13–14 February, pp. 1591–1597. ISBN 978-5-6042216-9-3. (In Russian).

3. Wadhawan A., Kumar P. Sign language recognition systems: a decade systematic literature review. *Archives of Computational Methods in Engineering*, 2021, vol. 28, pp. 785–813. DOI: 10.1007/s11831-019-09384-2.

4. Elakkiya R., Grif M.G., Prikhodko A.L., Bakaev M.A. Recognition of Russian and Indian sign languages used by the deaf people. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo*

[*] *Received 20 February 2021.*
Acknowledgments: The reported study was funded by RFBR, project number 19-57-45006.

tekhnicheskogo universiteta = Science bulletin of the Novosibirsk state technical university, 2020, no. 2–3 (79), pp. 57–76. DOI: 10.17212/1814-1196-2020-2-3-57-76.

5. Gruber I., Ryumin D., Hrúz M., Karpov A. Sign language numeral gestures recognition using convolutional neural network. *Interactive Collaborative Robotics*. Cham, Springer, 2018, pp. 70–77.

6. Konstantinov V.M., Orlova Yu.A., Rozaliev V.L. Razrabotka 3D-modeli tela cheloveka s ispol'zovaniem MS Kinect [The Development of 3D human body model using MS Kinect]. *Izvestiya Volgogradskogo gosudarstvennogo tekhnicheskogo universiteta = Izvestia of Volgograd State Technical University*, 2015, no. 6 (163), pp. 65–69.

7. Kagirov I., Ivanko D., Ryumin D., Axyonov A., Karpov A. TheRuSLan: Database of Russian Sign Language. *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, Marseille, 2020, pp. 6079–6085.

8. Mukushev M., Imashev A., Kimmelman V., Sandygulova A. Automatic classification of handshapes in Russian sign language. *Proceedings of the LREC 2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, Marseille, 2020, pp. 165–169.

9. Kimmelman V., Imashev A., Mukushev M., Sandygulova A. Eyebrow position in grammatical and emotional expressions in Kazakh-Russian Sign Language: A quantitative study. *Plos One*, 2020, vol. 15 (6). DOI: 10.1371/journal.pone.0233731.

10. Burkova S.I., Kimmel'man V.I., eds. *Vvedenie v lingvistiku zhestovykh yazykov. Russkii zhestovyi yazyk* [Introduction to the linguistics of sign languages. Russian sign language]. Novosibirsk, NSTU Publ., 2019. 356 p.

11. *Slovar' russkogo zhestovogo yazyka* [Dictionary of Russian Sign Language]. Available at: https://www.spreadthesign.com/ru (accessed 31.08.2021).

12. Bazoev V.Z. *Clovar' russkogo zhestovogo yazyka* [Dictionary of Russian Sign Language]. Moscow, Flinta Publ., 2009. 528 p.

13. *Supervisely*. *The leading platform for entire computer vision lifecycle*. Available at: https://supervise.ly/ (accessed 31.08.2021).

14. Kebets P.L. [Markers of the transition of the move in the discourse of Russian sign language]. *Materialy Sed'moi Konferentsii po tipologii i grammatike dlya molodykh issledovatelei* [Materials of the Seventh Conference on typology and grammar for young scholars], St. Petersburg, November 4–6, 2010, pp. 75–80. (In Russian).

15. System MediaPipe Pose. Available at: https://google.github.io/mediapipe/solutions/pose.html (accessed 31.08.2021).

16. Elakkiya R., Rajalakshmi E. ISLAN. *Mendeley Data*, 2021, ver. 1. DOI: 10.17632/rc349j45m5.1.

17. Dictionary of Indian Sign Language. Available at: http://indiansignlanguage.org/dictionary/ (accessed 31.08.2021).

18. Lugaresi C., Tang J., Nash H., McClanahan C., Uboweja E., Hays M., Zhang F., Chang C.-L., Yong M.G., Lee J., Chang W.-T., Hua W., Georg M., Grundmann M. *Mediapipe: A framework for building perception pipelines*. arXiv preprint arXiv: 1906.08172. 2019.

19. Le Q.V. Building high-level features using large scale unsupervised learning. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, 2013, pp. 8595–8598.

20. *Formal SignWriting (draft-slevinski-formal-signwriting-07)*. Available at: https://tools.ietf.org/id/draft-slevinski-formal-signwriting-07.html (accessed 31.08.2021).

Для цитирования:

Распознавание русского и индийского жестовых языков на основе машинного обучения / М.Г. Гриф, Р. Элаккия, А.Л. Приходько, М.А. Бакаев, Е. Раджалакшми // Системы анализа и обработки данных. – 2021. – № 3 (83). – С. 53–74. – DOI: 10.17212/2782-2001-2021-3-53-74.

For citation:

Grif M.G., Elakkiya R., Prikhodko A.L., Bakaev M.A., Rajalakshmi E. Raspoznavanie russkogo i indiiskogo zhestovykh yazykov na osnove mashinnogo obucheniya [Recognition of Russian and Indian Sign Languages based on machine learning]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2021, no. 3 (83), pp. 53–74. DOI: 10.17212/2782-2001-2021-3-53-74.