

УДК 004.5

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПРОГРАММНО-АППАРАТНЫХ СРЕДСТВ В ЗАДАЧАХ РАСПОЗНАВАНИЯ ЖЕСТОВОЙ РЕЧИ*

М.Г. ГРИФ¹, А.Н. КОЗЛОВ²

¹ 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, доктор технических наук, профессор, заведующий кафедрой автоматизированных систем управления. E-mail: grifmg@mail.ru

² 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, аспирант кафедры автоматизированных систем управления. E-mail: andrey.n.kozlov@gmail.ru

В данной статье рассмотрены особенности грамматики жестового русского языка, такие как морфология и синтаксис, после чего приведены отличия этой грамматики от грамматики разговорного и письменного русского языка. Далее представлен сравнительный анализ характеристик аппаратных и программных средств, которые могут быть использованы в задачах распознавания жестовой речи. Рассмотрены наиболее популярные виды устройств, такие как камеры, браслеты, перчатки и другие, которые применяются в области распознавания, а также подробно описаны наиболее популярные (их состав, возможности и отличия друг от друга), которые уже используются в области человеко-машинного взаимодействия в различных проектах. Приведены примеры данных, которые можно получить с этих устройств, а также то, как с ними можно работать. Далее рассмотрены некоторые наиболее успешные инструменты для работы с этими данными, такие как Knime, Weka, R-Studio, способы обработки видеопотока, а также упомянуты методы и алгоритмы, такие как HMM, ANN, SVM и другие, которые используются при решении задач машинного обучения и в работе с Big Data.

Ключевые слова: грамматика русского жестового языка, распознавание жестов, аппаратные и программные средства, задачи классификации, машинное обучение, большие данные, компьютерное зрение, сурдоперевод, человеко-компьютерное взаимодействие.

ВВЕДЕНИЕ

Сейчас в мире примерно 10 % населения любой страны мира – это лица с ограниченными возможностями здоровья по слуху. Большинство из них в повседневной жизни использует разговорный жестовый язык. Однако остро стоит проблема физической нехватки сурдопереводчиков для обеспечения библингвистических потребностей глухих людей во всех областях их жизнедеятельности. Поэтому во всем мире активно проводятся исследования в области

* Статья получена 18 июня 2014 г.

создания систем компьютерного сурдопереводчика. Так, например, такие компании, как Google, Microsoft и Intel, имеют отдельные лаборатории, где ведутся исследования в этой области.

Основная проблема распознавания жестовой речи состоит в том, что недостаточно распознавать отдельные жесты, как, например, при распознавании жестов в человеко-компьютерном взаимодействии. Необходимо распознавать непрерывную жестовую речь, ее смысл.

1. ПОСТАНОВКА ЗАДАЧИ

Перед тем как приступить к созданию архитектуры будущей системы и выбору аппаратных и программных средств, стоит разобраться, что же такое жестовая речь и чем она отличается от устной речи.

В статье рассмотрена грамматика жестовой речи, а именно морфология и синтаксис русского жестового языка; даны сравнительный анализ и краткая характеристика имеющихся перспективных аппаратных средств, а также представлены основные подходы, которые используются для решения данной проблемы.

2. ГРАМАТИКА ЖЕСТОВОЙ РЕЧИ

Рассмотрим особенности грамматики русского жестового языка (РЖЯ) морфологию и синтаксис.

«Морфология – это та часть грамматического строя языка, которая объединяет грамматические классы слов (части речи), принадлежащие этим классам грамматические (морфологические) категории и формы слов» [1].

В русском жестовом языке не принято разделять жесты на части речи. Так, один жест может обозначать ХОДИТЬ НА ЛЫЖАХ, ЛЫЖНИК, ЛЫЖИ [2].

С функциональной точки зрения в РЖЯ существуют категории имени и глагола, так как в нем с помощью жестов могут быть переданы неизменяемые во времени сущности и объекты и изменяющиеся во времени ситуации.

Анализ теоретических и эмпирических материалов позволил нам сделать вывод о том, что в русском жестовом языке выделяются классы жестов, тождественные частям речи в звучащем языке: имя существительное, имя прилагательное, имя числительное, местоимение, глагол, наречие, предлог, союз, частица, причастие, предикатив, междометие, модальные слова [3]. Следовательно, чтобы описать морфологию РЖЯ, необходимо описать по вышеупомянутым критериям выделенные группы жестов:

1) жесты-существительные – жесты, которые обозначают только неизменяемые во времени сущности, объекты, являющиеся участниками какого-либо действия;

2) жесты-прилагательные – жесты, используемые для обозначения непроецессуального признака. У этой группы жестов формально не выражаются категории числа и рода. Они всегда употребляются в форме, аналогичной форме именительного падежа единственного числа мужского рода;

3) жесты-числительные – жесты, которые используют для обозначения количества и порядка предметов при счете;

4) жесты-местоимения – жесты, которые используют для указания на лицо, предмет или признак;

5) жесты-глаголы – жесты, которые всегда используются для выражения ситуации, имеющейся во времени;

6) жесты-наречия – жесты, которые используют для обозначения признака действия, качества или предмета;

7) жесты-предикативы, т. е. жесты с модальными значениями долженствования, необходимости, возможности: должен, можно, надо, нельзя;

8) предлоги в РЖЯ являются средством выражения пространственных отношений. Словосочетание «люстра над столом» исполняется так: жест ЛЮСТРА над жестом СТОЛ.

Следующее, что стоит рассмотреть, это синтаксис русского жестового языка.

Термин «синтаксис» употребляется в языкознании в двух значениях:

1) существующая в языке система типов соединения слов в предложения и система предложений разных типов;

2) раздел грамматики, изучающий эту часть языковой системы [1].

Таким образом, «в синтаксис входят все явления связей слов и образования единиц, конструируемых на основе этих связей; законы строения простых и сложных предложений, правила их распространения и функционирования и правила вхождения предложения в текст [4].

В русском жестовом языке жест является аналогом слова. У данной единицы нет формы, аналогичной склоняемым и спрягаемым словоформам, а предложение в РЖЯ – это высказывание, выполняющее те же функции, что и предложение в устной речи.

По цели высказывания все предложения в РЖЯ делятся на два класса: вопросительные и невопросительные. Вопросительные предложения заключают в себе поиск сообщения: говорящий информирует о том, что он хочет получить информацию от другого (других). Невопросительные предложения традиционно делятся на три группы: предложения повествовательные, побудительные и предложения со значением желания [1].

Также предложения разделяются на простые и сложные.

Простое предложение – это такое высказывание, которое образовано по специально предназначенной для этого структурной схеме, обладает грамматическим значением предикативности и своей собственной семантической структурой.

Сложное предложение – это целостная синтаксическая единица, представляющая собой грамматически оформленное сочетание предложений и функционирующая в качестве сообщения о двух или более ситуациях и об отношениях между ними [1].

3. АППАРАТНЫЕ СРЕДСТВА

После того как мы рассмотрели основы грамматики русского жестового языка, перейдем к выбору средств для реализации системы.

Все устройства, которые используются для распознавания жестов, делятся на две группы: устройства, основанные на видеоданных, и устройства, использующие данные ускорения.

Первая группа устройств – это различного вида камеры, уровень распознавания которых зависит от их характеристик (разрешение сенсора, угол обзора и частота кадров). Но лучший результат дают системы, которые помимо RGB-сенсора имеют IR Depth-сенсор (сенсор глубины), который применяется для распознавания жестов на небольшом расстоянии.

Одним из примеров удачного технического решения является камера от компании Creative, которую сложно назвать просто камерой. Скорее это набор сенсоров. Она содержит два микрофона, позволяющих значительно улучшить качество распознавания речи, встроенную RGB и инфракрасную камеры [5].

Другой хороший пример – бесконтактный контроллер Kinect, разработанный фирмой Microsoft. Данное устройство имеет RGB-сенсор, инфракрасный излучатель и ИК-сенсор. Оба устройства имеют свои отличительные особенности, поэтому трудно сказать, какое из них может больше подходить для распознавания жестов. Так, например, Kinect применяется на дистанции от 1,8 до 3,0 метров, когда камера от Creative работает от 15 см до одного метра. Другим различием является то, что Creative больше подходит для распознавания жестов, идентификации и треккинга объектов, рук, лица, их синтеза, распознавания голоса, в то время как функционал Kinect направлен на идентификацию и распознавание жестов и тела человека, чтобы выделять его в окружающей среде дома [5, 6].

Еще одно устройство, которое относится к этой категории, но не является камерой, это Leap Motion. Leap Motion – это небольшое устройство размером с 3G-модем, несет в себе Monochromatic IR cameras – 2 шт. и Infrared LEDs – 3 шт. Leap Motion отслеживает все 10 пальцев с точностью до 1/100 мм. Устройство имеет угол обзора в 150° и ось Z для определения высоты [7]. Leap Motion появился на рынке позднее Kinect и Creative, но уже успел заинтересовать своими возможностями. Так, например, компании Asus и HP уже стали использовать его в своих ноутбуках [8].

Все три устройства (Kinect, Creative и Leap Motion) и их сенсоры изображены на рис. 1.



Рис. 1. Microsoft Kinect, Creative Interactive Gesture Camera, Leap Motion

Вторая группа устройств – устройства, основанные на использовании данных акселерометра, гироскопа и магнитометра. В основном данные устройства используются только для человеко-компьютерного взаимодействия. Например, игровая консоль Nintendo Wii, которая использует беспроводные контроллеры Wii Remote и Wii MotionPlus для отслеживания движений рук, и приставка PlayStation 3 с ее PlayStation Move. Данные устройства не способны различить движение кистей и пальцев, поэтому их рассматривать не будем.

В настоящее время на рынке не существует систем, которые, получая данные акселерометра и гироскопа, могли бы использоваться для сурдоперевода. Пока что это всего лишь прототипы, например кольца, перчатки с набором датчиков.

Устройство, которое должно решить проблему, это браслет MYO компании Thalmic Labs. Он появится в продаже уже во второй половине 2014 года. Браслет содержит в себе 3-осевой акселерометр, 3-осевой гироскоп, а также 3-осевой магнитометр, но основной особенностью устройства является наличие специального датчика, который получает информацию, считывая электрические импульсы в мышцах двигающейся руки. По словам Thalmic Labs, гаджет способен улавливать движение «каждого отдельного пальца», а также «небольшие движения руки в любом направлении». Гаджет запрограммирован на игнорирование случайных движений пользователя и предупреждает владельца при помощи тактильной обратной связи, если он активен и выполнил какую-либо команду [9].

4. ПРОГРАММНЫЕ СРЕДСТВА

Данные со всех вышеописанных устройств на выходе различны. Так, для устройств, которые основаны на потоке видеоданных, это набор статичных изображений, которые сменяют друг друга с определенной частотой. Для устройств, которые работают с акселерометром и гироскопом, это набор координат X , Y , Z . С данными акселерометра можно работать сразу, видеокadres необходимо преобразовать, а именно необходимо извлечь у объекта наблюдения элементарные составляющие. Самый простой способ выделить объект – это цвет. Для описания цвета используется два пространства: линейное и нелинейное. К линейному относится RGB (Red, Green, Blue – цветовая модель, описывающая способ синтеза цвета для цветовоспроизведения), а к нелинейному – HSV (Hue – тон, Saturation – насыщенность, Value – интенсивность).

Чаще всего для работы с изображениями используется библиотека OpenCV (Open Source Computer Vision) – библиотека алгоритмов компьютерного зрения, обработки изображений и численных алгоритмов общего назначения с открытым кодом. Реализована на C/C++ [10]. Помимо этой библиотеки имеются Intel Perceptual Computing SDK и Kinect SDK, которые позволяют работать с видеопотоком как на низком (OpenCV), так и на высоком уровне (используя готовые алгоритмы распознавания жестов, положения рук, лица, голоса).

После получения выделенного объекта или набора данных ускорения стоит решить задачу определения жеста. Данная задача является задачей классификации, и на сегодняшний день есть огромное количество методов и алгоритмов машинного обучения для ее решения. Так, например, байесовские сети, искусственные нейронные сети (ANN), деревья решений, методы опорных векторов (SVM) и другие.

Для того чтобы определить, какой из методов машинного обучения будет применен для решения задачи, используют следующие инструменты: KNIME, Weka, RapidMiner, Prediction, язык R. На данный момент это самые популярные инструменты анализа данных, с помощью которых можно проверить различные гипотезы, попробовать различные методы и алгоритмы на небольших наборах данных.

От выбранного метода будет зависеть точность решения задачи. Ключевыми можно выделить HMM, SVM и ANN алгоритмы. HMM (скрытая марковская модель) – статистическая модель, имитирующая работу процесса, похожего на марковский процесс с неизвестными параметрами. HMM является простейшей байесовской сетью доверия. Именно эта модель лежит в основе распознавания речи, также ее используют и для распознавания каких-либо образов, например, для распознавания лиц. SVM (метод опорных векторов) – набор схожих алгоритмов обучения с учителем, использующихся для задач классификации и регрессионного анализа. Суть метода – это перевод исход-

ных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости в этом пространстве. Для того чтобы разобраться в этом подходе, есть хороший пример – «коровы и волки» [11]. Простой пример SVM в среде KNIME изображен на рис. 2. ANN (искусственная нейронная сеть) – математическая модель, построенная по принципу биологических нейронных сетей, т. е. мозга. ANN представляет собой систему соединенных и взаимодействующих между собой простых процессоров (искусственных нейронов). Каждый процессор подобной сети имеет дело только с сигналами, которые он периодически получает, и сигналами, которые он периодически посылает другим процессорам. И тем не менее, будучи соединенными в достаточно большую сеть с управляемым взаимодействием, такие локально простые процессоры вместе способны выполнять довольно сложные задачи [12]. В отличие от других подходов, нейронные сети могут работать с чистыми данными без каких-либо гипотез второго уровня. Это связано с тем, что НС сама формирует гипотезы, которые могут находиться за пределами понимания. Но при использовании нейронных сетей приходится платить. Для того чтобы построить эти гипотезы, ей требуется огромное количество образцов, а ее обучение занимает очень длительное время. Так, например, чтобы обучить сеть распознавать мяч с картинки на обычном стационарном компьютере, потребуется около года.

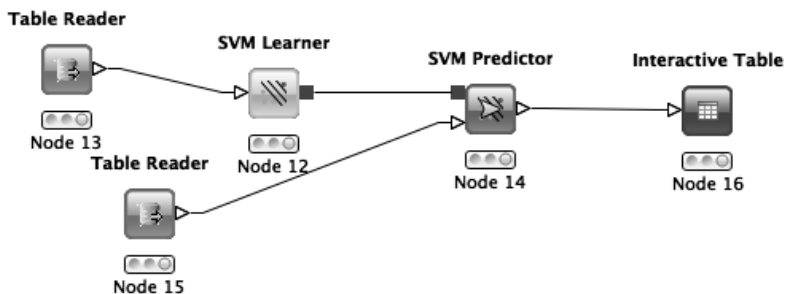


Рис. 2. Пример SVM в среде KNIME

ЗАКЛЮЧЕНИЕ

Мы рассмотрели грамматику жестовой речи русского языка, а также имеющиеся на данный момент средства – как аппаратные, так и программные. Пока не существует идеального решения, поэтому для более точных результатов лучше использовать комбинации из этих систем. Данная проблема является проблемой не только сурдоперевода, но также и проблемой человеко-машинного взаимодействия, а с быстрым развитием техники в ближайшее время появится еще больше технологий, которые можно будет использовать.

СПИСОК ЛИТЕРАТУРЫ

1. Русская грамматика: web-сайт. – URL: <http://rusgram.narod.ru/> (дата обращения: 15.12.2013).

2. *Зайцева Г.Л.* Жестовая речь. Дактилология: учебник для высших учебных заведений. – М.: ВЛАДОС, 2000. – 192 с.

3. *Королькова О.О.* Сопоставление морфологии русского звучащего и русского жестового языка // Научное творчество XXI века: материалы IV Всероссийской научно-практической конференции с международным участием, апрель 2011 г. – Красноярск, 2011. – С. 182–183. – (Приложение к журналу «В мире научных открытий»; вып. 1).

4. *Киммельман В.И.* Базовый порядок слов в русском жестовом языке: дипломная работа / Рос. гос. гуманитар. ун-т, Ин-т лингвистики. – М., 2010. – 163 с. – URL: <http://signlang.ru/science/read/kimmelman2/> (дата обращения: 10.12.2013).

5. Введение в естественно-интуитивное взаимодействие с компьютером: [учебный курс] / Ю. Березовская, В. Некрасова, К. Носов, О. Юфрякова; Север. (Аркт.) федер. ун-т им. М.В. Ломоносова // Национальный открытый университет «ИНТУИТ»: web-сайт. – Опубликовано: 23.10.2013. – URL: <http://www.intuit.ru/studies/courses/10619/1103/lecture/18218> (дата обращения: 15.12.2013).

6. Kinect – новый датчик для роботов? // RoboCraft: блог любителей робототехники, электроники и программирования. – URL: <http://robocraft.ru/blog/news/361.html> (дата обращения: 15.12.2013).

7. Leap motion teardown // SparkFun: retail online store of electronics. – URL: <https://learn.sparkfun.com/tutorials/leap-motion-teardown/the-outsidehtml> (дата обращения: 15.12.2013).

8. *Lee D.* Leap motion seals HP deal to embed gesture control technology // BBC News Technology. – Submitted on 16 April 2013. – URL: <http://www.bbc.co.uk/news/technology-22166424> (дата обращения: 16.12.2013).

9. *Кучинский Я.* Браслет MYO и мир Будущего // 3DNews – Daily Digital Digest: web-сайт. – URL: <http://www.3dnews.ru/645483> (дата обращения: 16.12.2013).

10. OpenCV. Компьютерное зрение // RoboCraft: блог любителей робототехники, электроники и программирования. – URL: <http://robocraft.ru/blog/computervision/2.html> (дата обращения: 16.12.2013).

11. *Lamp G.* Why use SVM? – URL: <http://www.yaksis.com/posts/why-use-svm.html> (дата обращения: 13.12.2013).

12. *Stergiou Ch., Siganos D.* Neural networks: report. – URL: http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html (дата обращения: 10.12.2013).

Гриф Михаил Геннадьевич – доктор технических наук, профессор кафедры автоматизированных систем управления, заведующий кафедрой автоматизированных систем управления Новосибирского государственного технического университета. Основное направление научных исследований – методы оптимального проектирования человеко-машинных систем. Имеет более 200 публикаций, в том числе 5 монографий. E-mail: grifmg@mail.ru.

Козлов Андрей Николаевич – аспирант кафедры автоматизированных систем управления Новосибирского государственного технического университета. Основное направление научных исследований – методы оптимального проектирования человеко-машинных систем. E-mail: andrey.n.kozlov@gmail.ru.

Approaches Of Recognition Sign Language*

M.G. Grif¹, A.N. Kozlov²

¹ *Novosibirsk State Technical University, 20 Karl Marks Avenue, Novosibirsk, 630073, Russian Federation, doctor of Technical Sciences, professor. E-mail: grifmg@mail.ru*

² *Novosibirsk State Technical University, 20 Karl Marks Avenue, Novosibirsk, 630073, Russian Federation, postgraduate student of the department of automated control systems. E-mail: andrey.n.kozlov@gmail.ru*

The features of the grammar of sign language its Russian morphology and syntax in the paper has been described. Differences between sign language and spoken and written Russian language has been shown. A comparative analysis of the characteristics of hardware and software that can be used in pattern recognition of sign language has been presented after. The most popular kinds of devices like cameras, bracelets and other, which are used in the field of pattern recognition has been considered, as well as detail the most popular (the possibilities and the composition), which are already used in the area of human-computer interaction has been described. Set of data that can be obtained with these devices has been shown. Some of the most successful tools for working with the data, the methods of processing the video stream, and also mention the methods like HMM, SVM, ANN, etc. that are used in the solutions of the problems of machine learning and working with Big Data has been considered.

Keywords: Russian colloquial sign language, gesture recognition, hardware and software, statistical classification, machine learning, Big Data, computer vision, human-computer interaction.

REFERENCES

1. *Russkaya grammatika*: web-sait [Russian grammar: website]. Available at: <http://rusgram.narod.ru/> (accessed 15.11.2013).
2. Zaitseva G.L. *Zhestovaya rech'. Daktilologiya* [Sign speech. Dactilology]. – Moscow, VLADOS Publ., 2000. 192 p.

* Received 18 June 2014.

3. Korol'kova O.O. [Comparison of morphology of the Russian sounding and Russian sign language]. *Nauchnoe tvorchestvo XXI veka: materialy IV Vserossiiskoi nauchno-prakticheskoi konferentsii s mezhdunarodnym uchastiem. Prilozhenie k zhurnalu «V mire nauchnykh otkrytii». Vypusk 1* [Scientific creativity of the XXI century. Proceedings of IV All-Russian scientific and practical conference with international participation. Supplement to the journal in the “World of scientific discoveries”], 2011, iss. 1, pp. 182–183. (In Russian)

4. Kimmel'man V.I. *Bazovyi poryadok slov v russkom zhestovom yazyke*. Diplomnaya rabota. [The basic word order in Russian sign language. Graduate work]. Moscow, 2010. 163 p. Available at: <http://signlang.ru/science/read/kimmelman2/> (accessed 10.12.2013).

5. Berezovskaya Yu., Nekrasova V., Nosov K., Yufryakova O. *Vvedenie v estestvenno-intuitivnoe vzaimodeistvie s komp'yuterom* [Introduction to natural intuitive interaction with the computer]. Available at: <http://www.intuit.ru/studies/courses/10619/1103/lecture/18218> (accessed 15.12.2013).

6. *Kinect – novyi datchik dlya robotov?* [Kinect – a new sensor for robots?]. Available at: <http://robocraft.ru/blog/news/361.html> (accessed 15.12.2013).

7. Leap motion teardown. Available at: <https://learn.sparkfun.com/tutorials/leap-motion-teardown/the-outsidehtml> (accessed 15.12.2013).

8. Lee D. Leap motion seals HP deal to embed gesture control technology. BBC News Technology. Submitted on 16 April, 2013. Available at: <http://www.bbc.co.uk/news/technology-22166424> (accessed 16.12.2013).

9. Kuchinskii Ya. *Braslet MYO i mir budushchego* [Bracelet MYO and world of the future]. Available at: <http://www.3dnews.ru/645483> (accessed 16.12.2013).

10. OpenCV. Computer vision. Available at: <http://robocraft.ru/blog/computer-vision/2.html> (accessed 16.12.2013).

11. Lamp G. Why use SVM? Available at: <http://www.yaksis.com/posts/why-use-svm.html> (accessed 13.12.2013).

12. *Stergiou Ch., Siganos D.* Neural networks: report. Available at: http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html (accessed 10.12.2013).