

**К. Ш. Курбанова**, главный специалист, e-mail: kemalewamil@gmail.com,  
Учебно-инновационный центр, Институт Информационных Технологий  
Министерства науки и образования Азербайджана, г. Баку, Азербайджан

### Исследование этапов, типов моделирования и методов распознавания жестов

*В современном обществе, где происходит быстрое развитие методов коммуникации между людьми, невозможно оставить в стороне проблему общения с внешним миром и социальной интеграции людей с ограниченными возможностями. Для людей с ограниченными слуховыми и речевыми возможностями жест является единственным методом общения. В исследовании представлена информация о статических и динамических жестах, которые позволяют людям с проблемами общения обмениваться информацией. Отмечено, что проблема автоматического распознавания жестов решается различными математическими методами, алгоритмами и компьютерными системами. Показаны преимущества и недостатки 2D- и 3D-моделей для последовательного распознавания жестов рук. Отмечено, что создание национальной электронной базы данных для решения задачи распознавания национального дактильного алфавита и жестов в Азербайджанской Республике является актуальной проблемой. Рассмотрены известные методы распознавания — метод искусственной нейронной сети, скрытый метод Маркова, метод случайного леса и метод обозначения жеста руки с помощью перчаток, раскрашенных маркерами разного цвета. Обсуждается создание национальной электронной базы данных для распознавания национального дактильного алфавита и жестов, частично устраняющее существующий барьер в образовании и общении людей.*

**Ключевые слова:** язык жестов, информационные технологии, моделирование жестов, методы распознавания жестов, алгоритм распознавания жестов

#### Введение

В настоящее время во всем мире очень важным и актуальным аспектом для взаимодействия людей является изучение иностранных языков. Оно позволяет реализовать процесс общения, а также создает возможность изучить культуру, искусство и историю других народов.

Из истории развития человечества известно, что первобытные люди использовали жесты для передачи информации, таким образом, первичным способом общения человека являются жесты. Жестовый язык является универсальным и дает возможность общаться с людьми независимо от их этнического происхождения, национального языка, возраста или ограничения в психологическом и физическом (глухонемые) развитии. Универсальность языка жестов находит применение в разных областях науки и промышленности, например при взаимодействии человека с роботом, в авиации (при объяснении пассажирам правил безопасности при экстремальных ситуациях), в жестах регулировщика дорожного движения и т. д.

Видно, что жесты широко используются в разных сферах деятельности человека, а для людей с ограниченными возможностями (слуховыми и речевыми) жест является единственным методом общения.

В современном обществе, где происходит быстрое развитие методов коммуникации между людьми, невозможно оставить в стороне проблему общения с внешним миром и социальной интеграции людей с ограниченными возможностями. Усовершенствование легких в управлении человеко-машинных интерфейсов для различных приложений является актуальной научной задачей.

#### Язык жестов и процесс распознавания жестов

Язык жестов — это невербальный способ коммуникации с помощью комбинаций жестов рук, а также сочетания мимики лица, формы глаз, губ и телодвижений [1].

Жесты могут быть статическими (одна конкретная позиция) или динамическими (последовательность позиций). Большинство челове-

ских жестов — это динамические жесты. Один динамический жест всегда состоит из нескольких элементов. Завершающим и самым важным этапом в распознавании жестов является их классификация. Классификацию динамических жестов можно вести только при их непосредственном наблюдении и отслеживании [2].

Математическая интерпретация человеческих движений вычислительным устройством дает возможность распознавания человеческого жеста.

Процесс распознавания жестов используется при вводе информации в электронное устройство в целях управления, идентификации или коммуникации. Распознавание жестов является системной задачей, включающей в себя моделирование и анализ движений, шаблоны распознавания и машинное обучение [3].

Система распознавания жестов — это совокупность компьютерных технологий и математических алгоритмов, которая позволяет решать задачу распознавания определенной группы жестов. При исследовании системы распознавания жестов необходимо рассмотреть обобщенную модель обработки информации, включающую в себя четыре этапа [4]:

- получение информации. Реализуется с помощью камеры, датчиков, перчатки и др.;
- моделирование жеста. Выполняется программой для анализа информации, сегментации руки, фильтрации шума, определения контуров и нормализации;
- извлечение признаков жестов и принятие решения;
- распознавание жестов и ответные действия.

На сегодняшний день разработаны разные системы моделирования распознавания языка жестов для различных языков, которые создают возможность упрощения в общении для глухонемых людей.

### Ручное моделирование

Взаимодействие человека с компьютером — одна из областей, где успешно применяется процесс распознавания жестов. Чтобы компьютерное распознавание было точным, жесты рук должны быть четко смоделированы. В системах распознавания жестов существуют два типа моделирования [5].

В 2D-моделировании изображение снимается с одной камеры и является единственным источником для классификации жестов. Для получения необходимой информации используется алгоритм сегментации.

В этом типе моделирования используются цветные перчатки или цветные маркеры для упрощения процесса распознавания.

В 3D-моделировании жест или знак рисуется в трехмерном пространстве. Использование дополнительных координат устраняет неопределенность изображения и обеспечивает большую точность распознавания. Однако увеличение объема данных приводит к увеличению нагрузки на процессоры в процессе распознавания.

Технология распознавания жестов, основанная на анализе трехмерной модели руки, реализуется на основе обработки одной или нескольких деталей конфигурации объекта — направления, положения ладони и основных точек пальцев в трехмерном пространстве. В этом подходе в качестве геометрических объектов используются направления пальцев руки, кончики и контуры пальцев. Трехмерная модель руки строится на основе свободного движения суставов. В зависимости от вида движения эти показатели могут совпадать или отличаться.

Визуальные методы распознавания жестов можно разделить на перечисленные ниже основные категории [6]:

- первая категория — это метод анализа данных входного изображения руки для восстановления полной модели кисти с 27 степенями свободы. Степени свободы определяют позиции и ориентации реального объекта. Моделирование руки представляет собой сложную задачу, поэтому используется упрощенная модель руки с 27 степенями свободы (рис. 1, а). Заранее созданные базы данных всех известных конфигураций руки и их визуализации используются в системе отслеживания руки в пространстве на основе 3D-модели руки;

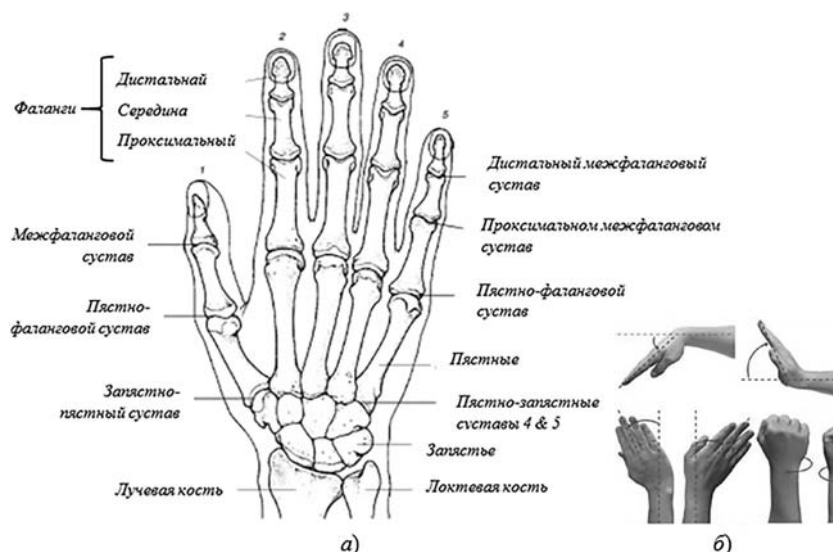


Рис. 1. Структура руки человека: а — модель руки с 27 степенями свободы; б — геометрические признаки

- вторая категория — это статистические методы, которые вместо восстановления полной модели руки предлагают построение признакового описания входного изображения и дальнейшую классификацию жестов. Основная задача в этом случае — это выбор подходящих характеристических признаков (рис. 1, б): геометрических признаков, таких как положение кончиков пальцев, контур руки, направление пальцев; не геометрических признаков, такие как цвет кожи, форма, текстура и др.;
- третья категория — это метрические методы распознавания жестов, которые предполагают построение некоторой метрики на множестве входных изображений и выполнение классификации за счет сравнения входного изображения с набором эталонов.

2D-изображения дают достаточную информацию для многих алгоритмов распознавания, для уточнения требуются и другие данные об объекте, такие как ориентация руки, форма руки и даже положение руки в пространстве. Чтобы получить эти данные, были разработаны методы трекинга и распознавания в пространстве. Для получения дополнительной информации применяется трехмерное кинематическое моделирование руки (3D-моделирование). При проецировании данной модели на плоскость получается двумерное изображение руки, которое сравнивается с поступающим изображением видеокadra. В результате сравнения распознается поза руки. Данный подход решает одновременно задачи обнаружения руки и распознавания позы руки (рис. 2) [7].

Рука человека обладает сочлененной структурой, и в зависимости от местоположения наблюдателя относительно исследуемого объекта возникают трудности в описании ее ориентации и формы. Предварительно требуется создание базы данных всех известных конфигураций руки и их визуализации (рендеринг), где каждая конфигурация содержит множество углов наклона суставов и параметры позы руки для описания состояния руки [8].

Функциональная схема метода видеоанализа для распознавания жестов представлена на рис. 3. Входными видеоданными являются цветной видеопоток и карта глубины, получаемые от сенсора Kinect v2. В случае недоступности сенсора Kinect v2 работа автоматически прерывается. На данном этапе остановка возможна в следующих случаях: возникновение ошибки как при получении цветных видеокadров, так и карты глубины; завер-

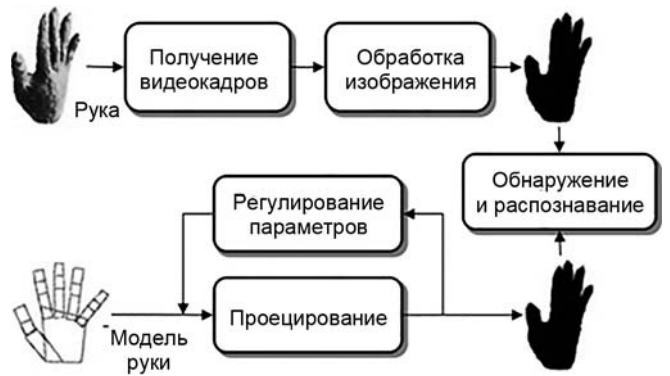


Рис. 2. Принципиальная схема подхода к обнаружению и распознаванию на основе 3D-модели рук

шение одного из описанных видеопотоков. На каждом 3D-кадре карты глубины происходит поиск людей на расстоянии от 1,2 до 3,5 м и вычисление 25-точечных 3D-моделей всех найденных скелетов людей. По оси  $Z$  трехмерного пространства определяется ближайшая скелетная модель и устанавливается слежение за ней. С помощью SDK Kinect v2 3D-координаты преобразуются в 2D-координаты (рис. 3) [9].

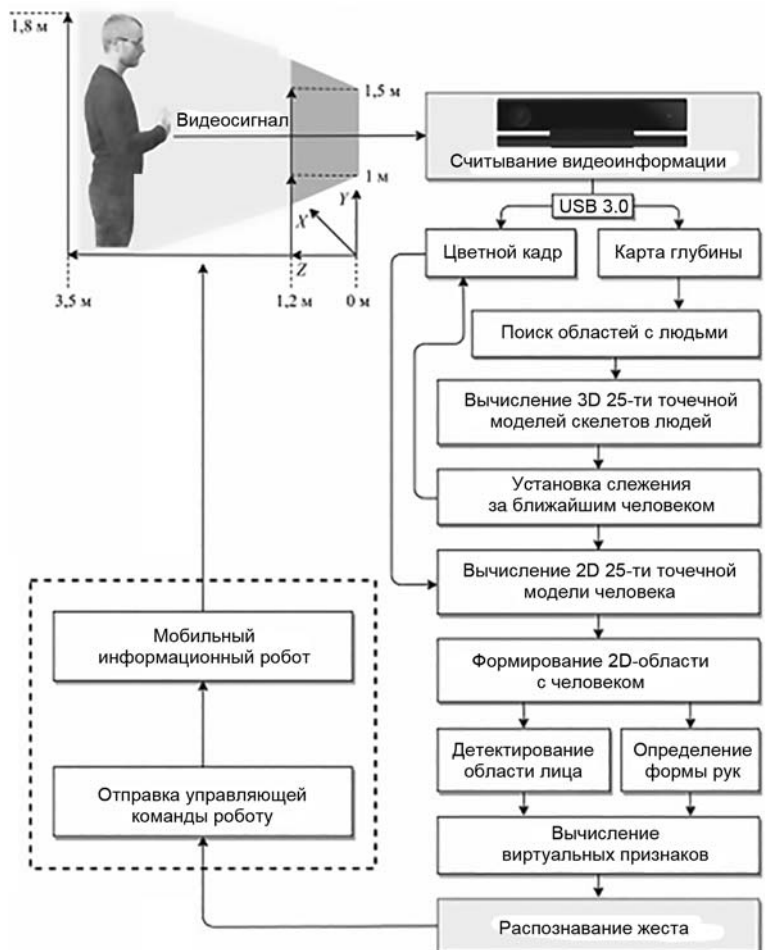


Рис. 3. Функциональная схема метода видеоанализа движений рук для распознавания жестов жестового языка

## Методы распознавания жестов

Методы, направленные на анализ особенностей жеста, основаны на реализации процесса распознавания жеста по показателям внешней формы и положения руки. Устройство, которое отслеживает и записывает жесты, функционирует правильно только в том случае, если силуэт пальцев четкий и точный. Полученные с использованием 2D-, 3D-камер и платформы Kinect изображения анализируются различными методами. Рассмотрим известные методы этой категории.

*Метод на основе искусственной нейронной сети (ИНС).* Метод на основе ИНС использует математическую модель, основанную на принципах организации и работы биологических нейронных сетей (рис. 4, см. третью сторону обложки). В задачах распознавания жестов характеристики жеста нормализуются и передаются на вход ИНС, на выходе сети получается жест. Здесь данные используются как инструмент машинного обучения. ИНС различаются по структуре и методам обучения. В проблемах распознавания наиболее часто используемые типы ИНС — это рекуррентные нейронные сети, независимо созданная карта Кохонена, искусственная нейронная сеть Эльмана и другие. Достоинством метода на основе ИНС является то, что он обнаруживает сложные нелинейные зависимости между переменными. Недостаток данного метода состоит в невозможности обучения при недостаточном объеме входных данных [10].

*Скрытая модель Маркова (СММ).* Математический аппарат СММ — это универсальный инструмент для распознавания жестов. СММ является очень мощным статистическим методом классификации наблюдаемых последовательностей данных. Для применения СММ в реальных приложениях необходимо решить три основных задачи: задачу оценки, задачу декодирования и задачу обучения. С помощью СММ можно распознавать как динамические, так и статические жесты, описываемые последовательными движениями рук. Достоинством СММ является гибкость обучения, прозрачность модели [11].

*Метод на основе алгоритма случайного леса (АСЛ).* АСЛ — это алгоритм машинного обучения, используемый в задачах классификации, регрессии и кластеризации. Этот алгоритм используется для классификации конфигурации руки в задачах распознавания жестов. Эта технология была представлена Microsoft в 2010 г. Предлагаемый метод на основе дистанционной визуализации позволяет распознать положение ладоней обеих рук. Идентификация точек, которые играют ключевую роль в распознавании жестов, описываемых сотнями тысяч поз тела, осуществляется с помощью АСЛ. Алгоритм распознавания включает следующие шаги [12–14]: выделение изображения человека из общего изображения; разделение изображения человека на части (голова, конечности, тело); разделение выделенных частей на кластеры; определение положения интересующих частей для распознавания жестов на основе полученных кластеров.

*Метод распознавания жеста руки с помощью перчаток, раскрашенных цветами разного цвета.* Этот метод основан на использовании специального оборудования. Датчики, прикрепленные к пальцам, фиксируют движение руки и передают информацию на компьютер. На основе введенной информации создается трехмерная модель руки, полностью описывающая жест, и эта модель анимируется для отображения движения руки. Этот метод значительно упрощает процесс распознавания жестов. На изображении, где рука является доминирующим объектом, фон должен быть одного цвета. Недостатком этого метода является использование в перчатке близких друг к другу цветных маркеров. Алгоритм распознавания состоит из следующих шагов: описание руки; получение растрового изображения руки; сравнение с шаблонами баз данных; идентификация жеста (рис. 5) [15].

## Заключение

В Азербайджане принимаются соответствующие меры по созданию условий для адапта-

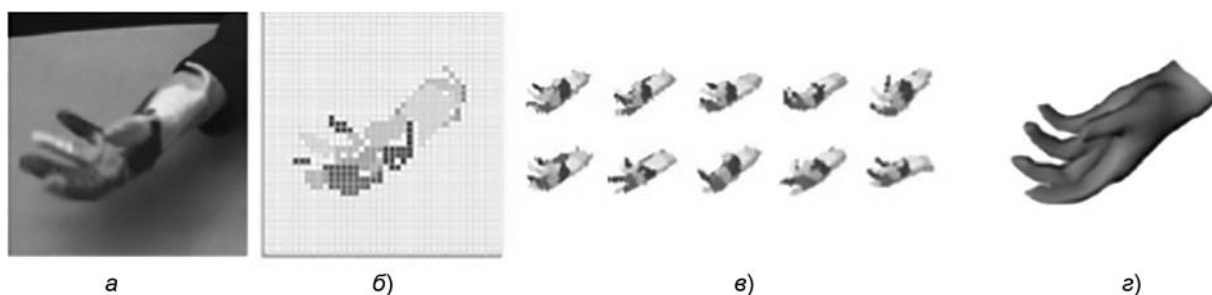


Рис. 5. Описание руки (а); растровый вид изображения руки (б); сравнение с шаблонами баз данных (в); жест назначается (г)

ции людей с ограниченными возможностями, в том числе людей с нарушениями слуха, а также для расширения их возможностей в области образования и трудоустройства. Азербайджан присоединился к Конвенции ООН "О правах инвалидов" в 2008 г.

Концепция развития "Азербайджан 2020: видение будущего", утвержденная в 2012 г., подчеркивает предоставление в стране качественных услуг здравоохранения и образования в области развития человеческого капитала и доступность этих услуг для различных социальных групп, включая людей с ограниченными возможностями. Одна из основных целей концепции — предотвратить изоляцию от общества людей с ограниченными возможностями [16].

Проект "Azeri Dactyl Alphabet" был разработан в Азербайджане для облегчения общения и интеграции людей с нарушениями речи и слуха в обществе. Программу, распознающую буквы дактильного алфавита и выражения жестов, поддерживают операционные системы Android и iOS. Эта система распознавания основана на технологии искусственного интеллекта [17].

Для более быстрого и легкого ввода информации из словаря жестов пользуются известной системой SignWriting, которая позволяет символизировать мимику лица, движения рук и тела. Эта система, состоящая из более чем тысячи символов, отличается однозначным написанием и частым использованием жестов. В последние годы наиболее популярной стала немецкая система обозначений HamNoSys (Hamberg Notation System). Это простая в использовании система кодирует жесты в системе Unicode и может регистрировать практически любой язык жестов в мире [18].

Компьютерные технологии помогают преодолеть барьер для социализации людей с нарушениями слуха и речи. Динамичное развитие в области искусственного интеллекта открывает различные пути создания системы распознавания жестового языка, которая выступает в роли своего рода переводчика для пользователей этого языка.

В результате быстрого развития и широкого применения информационных и коммуникационных технологий происходят радикальные изменения во всех сферах жизни общества. Расширяется общение людей, большими темпами развивается глобализация в экономической и социально-культурной сферах, возрастает роль информации во всех областях. Развитие компьютеров, интернета и мобильных телефонов с разнообразным программным обеспечением позволяет получать и передавать информацию быстрее и полнее за короткий период времени, а также увеличивает спрос на визуальную информацию.

Обмен информацией с помощью сети Интернет (IP-телефония, дистанционное обучение, интернет-магазины, электронная коммерция, чаты и т. д.), является нормой повседневной жизни. Одним из основных требований научно-технического общества является разработка новых методов и систем, преобразующих жесты в голос или наоборот для решения проблем социализации людей с ограниченным слухом. Создание национальной электронной базы данных, распознающей национальный дактильный алфавит и жесты, частично устранил существующий барьер в образовании и проблемах общения людей, проживающих в нашей республике.

#### Список литературы

1. **Дворина Н. Г.** Использование интерактивной компьютерной технологии распознавания жестов и речи на практических занятиях по иностранному языку // Интернет-журнал "Науковедение". 2015. № 3(7). С. 1–15. URL: <http://naukovedenie.ru/> (дата обращения: 11.12.2022).
2. **Hongyi L., Lihui W.** Распознавание жестов для взаимодействия с ИИ: от теории к последним достижениям. URL: <https://integral-russia.ru/2021/07/02/raspoznavanie-zhestov-dlya-vzaimodejstviya-s-ii-ot-teorii-k-poslednim-dostizheniyam/> (дата обращения: 18.01.2023).
3. **Сіряк Р. В.** Технологии идентификации и распознавания жестов // Вестник Восточноевропейского национального университета имени Владимира Даля. 2017. № 8 (238). С. 79–85.
4. **Parasuraman R., Sheridan T. B., Wickens C. D.** A model for types and levels of human interaction with automation, Systems, Man and Cybernetics, Part A: Systems and Humans // IEEE Transactions on. 2000. № 3(30). P. 286–297.
5. **Еникеев Д. Г., Мустафина С. А.** Использование контроллера leap motion для прикладных систем распознавания языка жестов // Информационные технологии. проблемы и решения. 2019. № 1(6). С. 50–54.
6. **Розалиев В. Л., Агафонов Г. В., Кириченко М. И.** Автоматизированное выделение кистей рук человека для распознавания жестовой речи // V междунар. науч.-техн. конф. Открытые семантические технологии проектирования интеллектуальных систем. 2015. 19–21 фев. С. 565–570.
7. **Wu Y., Huang T. S.** Vision-based gesture recognition: A review // Gesture-Based Communication in Human-Computer Interaction: International Gesture Workshop, GW'99 Gif-sur-Yvette, France, March 17–19, 1999 Proceedings. 1999. С. 103–115.
8. **Tomasi C., Petrov S., Sastry A.** 3D tracking = classification + interpolation // Proc. Ninth IEEE International Conference on Computer Vision (Nice, France, 2003). IEEE Computer Society. 2003. P. 1441–1448.
9. **Рюмин Д.** Метод автоматического видеоанализа движений рук и распознавания жестов в человеко-машинных интерфейсах // Научно-технический вестник информационных технологий, механики и оптики. 2020. № 4(20). С. 525–531.
10. **Murakami K., Taguchi H.** Gesture Recognition using Recurrent Neural Networks // ACM Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology (CHI '91). 1999. P. 237–242.
11. **Огнев И. В., Парамонов П. А.** Распознавание речи методами скрытых марковских моделей в ассоциативной осцилляторной среде // Технические науки. Информатика, вычислительная техника. 2013. № 3 (27). С. 115–126.

12. **Bobick A., Davis J.** An appearance-based representation of action // International Conference on Pattern Recognition. 1996. P. 307–312.

13. **Shotton J., Fitzgibbon A., Cook M., Sharp T., Finocchio M., Moore R., Kipman A., Blake A.** Real-Time Human Pose Recognition in Parts from Single Depth Images // In Proc. CVPR. 2011. P.1297–1304.

14. **Pugeault N., Bowden R.** Spelling It Out: Real-Time ASL Fingerspelling Recognition // In Proceedings of the 1st IEEE Workshop on Consumer Depth Cameras for Computer Vision, jointly with ICCV'2011. 2011. P. 1114–1119.

15. **Wang R. Y., Popovi J.** C Real-time hand-tracking with a color glove // ACM Trans. Graph. 2009. № 3(28). P. 1–8.

16. **Национальная стратегия развития информационно-го общества в Азербайджанской Республике на 2014–2020 годы.** URL: <http://president.az/articles/11312> (дата обращения: 11.02.2023).

17. **Проекты**, подготовленные под руководством сотрудников института, заняли первое и второе места в конкурсе "Ученые завтрашнего дня", URL: <https://ict.az/az/news/5352> (дата обращения: 15.10.2022).

18. **Кривонос Ю. Г., Крак Ю. В., Бармак А. В., Шкильнюк Д. В.** Конструирование и идентификация элементов жестовой коммуникации // Кибернетика и системный анализ. 2013. № 2(49). С. 3–14.

**K. Sh. Gurbanova**, Chief Specialist, e-mail: kemalewamil@gmail.com,  
Training-Innovation Centre, Institute of Information Technology  
of The Ministry of Science and Education of the Azerbaijan Republic, Baku, Azerbaijan

## Research of Stages, Types of Modeling and Methods of Gesture Recognition

*It is impossible to leave aside the problem of communication with the outside world and the social integration of people with disabilities in modern society, where there is a rapid development of communication methods between people. Gesture is the only method of communication for people with disabilities (auditory and language). The research provides information on static and dynamic gestures that allow people with communication problems to exchange information. It is noted that the problem of automatic gesture recognition is solved by various mathematical methods, algorithms and computer systems. The advantages and disadvantages of 2D and 3D models for sequential recognition of hand gestures are shown. It is noted that the creation of a national electronic database that recognizes the national dactyl alphabet and gestures in the Republic of Azerbaijan is an urgent problem. The well-known methods of this category are shown. Artificial neural network method, Hidden Markov method, Random Forest method and method of designating a hand gesture using gloves painted with markers of different colors. The creation of a national electronic database that recognizes the national dactyl alphabet and gestures is proposed, which will partially eliminate the existing barrier in education and problems of communication between people.*

**Keywords:** sign language, information technology, gesture modeling, gesture recognition methods, gesture recognition algorithm

DOI: 10.17587/it.30.85-90

### References

1. **Dvorina N. G.** The use of interactive computer technology for recognition of gestures and speech in practical classes in a foreign language, *Internet journal "Science Studies"*, 2015, no. 3(7), pp.1–15, available at: <http://naukovedenie.ru/> (date of access: 12/11/2022) (in Russian).

2. **Hongyi L., Lihui W.** Gesture recognition for AI interaction: from theory to recent advances, available at: <https://integral-russia.ru/2021/07/02/raspoznavanie-zhestov-dlya-vzaimodejstviya-s-ii-ot-teorii-k-poslednim-dostizheniyam/> (date of access: 01/18/2023).

3. **Siryak R. V.** Technologies for identification and recognition of gestures // Bulletin of the East Ukrainian National University named after Vladimir Dali, 2017, no. 8 (238), pp.79–85.

4. **Parasuraman R., Sheridan T. B., Wickens C. D.** A model for types and levels of human interaction with automation, Systems, Man and Cybernetics, Part A: Systems and Humans, *IEEE Transactions on*, 2000, no. 3 (30), pp. 286–297.

5. **Enikeev D. G., Mustafina S. A.** Using the leap motion controller for applied systems of sign language recognition, *Information technologies. problems and solutions*, 2019, no. 1(6), pp. 50–54 (in Russian).

6. **Rozaliev V. L., Agafonov G. V., Kirichenko M. I.** Automated selection of human hands for recognition of sign speech, *V Intern. sci.-tech. conf. Open semantic technologies for designing intelligent systems*, Feb 19–21, 2015, pp. 565–570 (in Russian).

7. **Wu Y., Huang T. S.** Vision-based gesture recognition: A review, *Gesture-Based Communication in Human-Computer Interaction: International Gesture Workshop, GW'99 Gif-sur-Yvette, France, March 17-19, 1999 Proceedings*, 1999, pp. 103–115.

8. **Tomasi C., Petrov S., Sastry A.** 3D tracking = classification + interpolation, *Proc. Ninth IEEE International Conference on Computer Vision (Nice, France, 2003)*, IEEE Computer Society, 2003, pp. 1441–1448.

9. **Ryumin D.** Method of automatic video analysis of hand movements and gesture recognition in human-machine interfaces, *Scientific and technical bulletin of information technologies, mechanics and optics*, 2020, no. 4(20), pp. 525–531 (in Russian).

10. **Murakami K., Taguchi H.** Gesture Recognition using Recurrent Neural Networks, *ACM Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology (CHI '91)*, 1999, pp. 237–242.

11. **Ognev I. V., Paramonov P. A.** Speech recognition by methods of hidden Markov models in an associative oscillatory environment, *Technical sciences. Informatics, computer technology*, 2013, no. 3 (27), pp. 115–126.

12. **Bobick A., Davis J.** An appearance-based representation of action, *International Conference on Pattern Recognition*, 1996, pp. 307–312.

13. **Shotton J., Fitzgibbon A., Cook M., Sharp T., Finocchio M., Moore R., Kipman A., Blake A.** Real-Time Human Pose Recognition in Parts from Single Depth Images, *In Proc. CVPR*, 2011, pp. 1297–1304.

14. **Pugeault N., Bowden R.** Spelling It Out: Real-Time ASL Fingerspelling Recognition, *In Proceedings of the 1st IEEE Workshop on Consumer Depth Cameras for Computer Vision, jointly with ICCV'2011*, 2011, pp. 1114–1119.

15. **Wang R. Y., Popovi J.** C Real-time hand-tracking with a color glove, *ACM Trans. graph.*, 2009, no. 3 (28), pp. 1–8.

16. **National strategy for the development of the information society in the Republic of Azerbaijan for 2014–2020**, available at: <http://president.az/articles/11312> (date of access: 11.02.2023).

17. **Projects** prepared under the guidance of the Institute's staff took first and second places in the "Scientists of Tomorrow" competition, available at: <https://ict.az/az/news/5352> (date of access: 10/15/2022).

18. **Krivosos Yu. G., Kрак Yu. V., Barmaк A. V., Shkilnyuk D. V.** Design and identification of elements of gesture communication, *Cybernetics and system analysis*, 2013, no. 2(49), pp. 3–14 (in Russian).