

УДК 004.93'1:81'221.24; 004.8; 004.85

Федоров К.Е.,
магистрант, 2 курс,
факультет автоматики и вычислительной техники, новосибирский
государственный университет, г. Новосибирск
Научный руководитель: Бакаев М.А.,
к.т.н., доцент

АНАЛИЗ МЕТОДОВ ПРЕДСТАВЛЕНИЯ ЖЕСТОВ ДЛЯ СИСТЕМ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Аннотация: В статье рассматриваются современные методы представления и распознавания жестов в системах компьютерного зрения. Проведен анализ подходов, основанных на сверточных нейронных сетях, рекуррентных архитектурах и методах выделения ключевых точек кисти руки. Особое внимание уделено использованию моделей MediaPipe Hands и LSTM-сетей для распознавания динамических жестов в режиме реального времени. Рассмотрены существующие датасеты, методы сегментации изображения и извлечения признаков жеста. Выполнено сравнение 2D- и 3D-подходов к анализу движений рук. Показано, что комбинирование CNN и LSTM позволяет эффективно выделять пространственно-временные признаки жестов и повышать точность распознавания. Результаты исследования могут быть использованы при разработке интеллектуальных интерфейсов взаимодействия человека и компьютера, а также систем автоматического распознавания жестового языка.

Ключевые слова: компьютерное зрение, распознавание жестов, жестовый язык, нейронные сети, MediaPipe, машинное обучение.

ANALYSIS OF GESTURE REPRESENTATION METHODS FOR COMPUTER VISION SYSTEMS

Научное издательство «Фокус знаний»

Summary: *The article discusses modern methods of gesture representation and recognition in computer vision systems. Approaches based on convolutional neural networks, recurrent architectures, and hand landmark detection are analyzed. Particular attention is paid to the use of MediaPipe Hands and LSTM models for real-time dynamic gesture recognition. Existing datasets, image segmentation methods, and gesture feature extraction approaches are reviewed. A comparison of 2D and 3D hand motion analysis methods is presented. The study shows that combining CNN and LSTM architectures improves the extraction of spatial-temporal gesture features and increases recognition accuracy.*

Keywords: *computer vision, gesture recognition, sign language, neural networks, MediaPipe, machine learning.*

Введение

Современные системы компьютерного зрения активно применяются в задачах анализа жестов и распознавания жестового языка. Подобные технологии используются в интеллектуальных интерфейсах, системах дополненной реальности, образовательных платформах и средствах коммуникации для людей с нарушениями слуха [1].

Одной из основных задач в данной области является выделение информативных признаков движения рук и их последующая классификация. Развитие глубокого обучения значительно повысило качество распознавания жестов благодаря использованию сверточных и рекуррентных нейронных сетей [2].

Несмотря на высокую точность современных моделей, проблема распознавания динамических жестов в реальном времени остается актуальной. Основные сложности связаны с изменением освещения, фона, скоростью движения рук и необходимостью обработки временной последовательности кадров.

Целью данной работы является анализ современных методов представления жестов и исследование архитектур нейронных сетей, применяемых в системах компьютерного зрения.

Методы исследования

Представление жестов в системах компьютерного зрения

Большинство современных систем распознавания жестов используют два основных подхода:

- анализ изображений RGB;
- анализ ключевых точек руки.

В первом случае нейронная сеть обрабатывает исходное изображение полностью. Во втором случае используются координаты суставов кисти руки, определяемые специальными моделями детекции.

Одним из наиболее распространенных инструментов является MediaPipe Hands, позволяющий выделять 21 ключевую точку кисти руки в реальном времени [3].

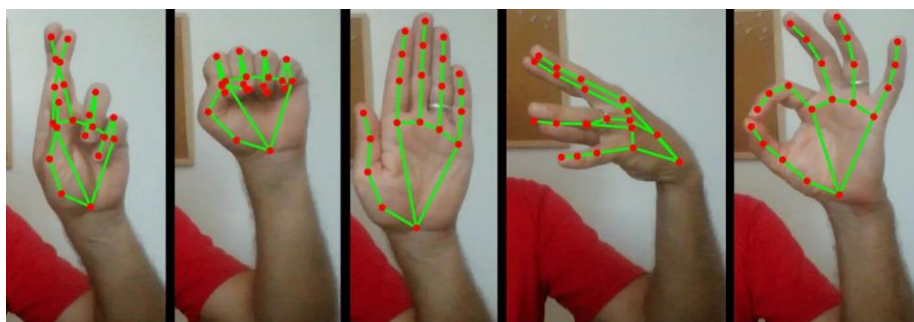


Рисунок 1. Пример определения ключевых точек кисти руки с использованием MediaPipe Hands

Использование ключевых точек значительно уменьшает объем входных данных и повышает устойчивость модели к изменению фона и освещения.

Сверточные нейронные сети

Сверточные нейронные сети (CNN) применяются для выделения пространственных признаков изображения. Такие модели эффективно определяют форму руки, положение пальцев и контуры жеста.

Наиболее распространенными архитектурами являются:

- YOLO;
- ResNet;
- EfficientNet;
- MobileNet.

Сверточные сети обеспечивают высокую точность при распознавании статических жестов, однако недостаточно эффективно работают с временной динамикой движений.

Рекуррентные нейронные сети и LSTM

Для анализа последовательности кадров используются рекуррентные сети и архитектуры Long Short-Term Memory (LSTM). Они позволяют учитывать временные зависимости между движениями рук.

LSTM-сети широко применяются при распознавании динамических жестов и жестового языка [4]. Входными данными для модели обычно являются последовательности координат ключевых точек руки.



Рисунок 2. Архитектура системы распознавания жестов на основе MediaPipe и LSTM

Использование LSTM позволяет анализировать траекторию движения кисти и распознавать сложные последовательности жестов.

Результаты исследования

В ходе анализа научных публикаций было установлено, что наиболее эффективными являются гибридные модели, объединяющие CNN и LSTM.

Таблица 1.

Сравнение методов распознавания жестов

Метод	Преимущества	Недостатки
CNN	Высокая точность распознавания формы руки	Не учитывает временную динамику
LSTM	Анализ последовательности движений	Требуется большого объема данных
MediaPipe + LSTM	Работа в реальном времени	Чувствительность к качеству камеры
3D CNN	Анализ пространственно-временных признаков	Высокая вычислительная сложность

Исследования показывают, что использование координат ключевых точек кисти позволяет существенно сократить вычислительные затраты по сравнению с обработкой полного изображения [5].

Дополнительно было установлено, что качество распознавания зависит от следующих факторов:

- объем обучающего датасета;
- качество разметки;
- разнообразие жестов;
- условия освещения;
- качество видеокамеры.

Наибольшую точность показывают модели, использующие одновременно пространственные и временные признаки.

Современные системы распознавания жестов демонстрируют высокую эффективность при использовании глубокого обучения. Однако существуют ограничения, связанные с необходимостью больших объемов размеченных данных и высокой вычислительной нагрузкой.

Дополнительной проблемой является ограниченное количество датасетов для русского жестового языка. Большинство существующих наборов данных ориентированы на английский или американский жестовый язык.

Перспективным направлением является использование мультимодальных систем, объединяющих анализ изображения, глубины сцены и траектории движения рук.

Выводы

В результате исследования были рассмотрены современные методы представления жестов в системах компьютерного зрения.

Установлено, что:

- модели MediaPipe позволяют эффективно выделять ключевые точки кисти руки;
- CNN-сети хорошо подходят для анализа пространственных признаков;
- LSTM-сети эффективны при обработке динамических жестов;
- гибридные архитектуры CNN + LSTM показывают наиболее высокую точность распознавания.

Полученные результаты могут быть использованы при разработке систем автоматического распознавания жестового языка и интеллектуальных интерфейсов взаимодействия человека и компьютера.

Литература:

1. Рюмин Д. А., Кагиров И. А., Аксёнов А. А., Карпов А. А. Аналитический обзор моделей и методов автоматического распознавания жестов и жестовых языков // Информационно-управляющие системы. 2021. № 6. С. 10–20 (дата обращения 01.05.2026).

2. Rastgoo R., Kiani K., Escalera S. Sign language recognition: A deep survey // Expert Systems with Applications. 2021. Vol. 164 (дата обращения 01.05.2026).
3. MediaPipe Hands. Google Research. [Электронный ресурс]. URL: <https://developers.google.com/mediapipe> (дата обращения 01.05.2026).
4. Koller O. Quantitative survey of the state of the art in sign language recognition // arXiv preprint arXiv:2008.09918. 2020 (дата обращения 01.05.2026).
5. Devineau G., Moutarde F., Xi W., Yang J. Deep learning for hand gesture recognition on skeletal data // Proceedings of FG 2018 (дата обращения 01.05.2026).