

# Метод ручной разметки данных для автоматического распознавания, классификации и прогнозирования действий по видеоданным

С. Е. Ильин

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

s.ilin2086@gmail.com

**Аннотация.** В работе формулируется метод ручной разметки данных для распознавания, классификации и прогнозирования действий по видеоданным, основанный на принципах объяснимого искусственного интеллекта и структурном подходе к определению действий. Приводится пример использования предлагаемого метода в условиях разметки видеоданных для распознавания действий человека. Производится сравнение предлагаемого метода с альтернативным вариантом разметки при решении соответствующей задачи.

**Ключевые слова:** автоматический анализ поведения; распознавание действий; классификация действий; прогнозирование действий; разметка данных; разметка видеоданных; ручная разметка данных

## I. ВВЕДЕНИЕ

На актуальном этапе развития технологий сбор и анализ видеоданных относятся к числу ключевых направлений информационной деятельности в глобальном масштабе. Популяризация портативных электронных устройств, широкое использование видеокamer, распространение интернета вещей, а также рост количества цифровых платформ для удаленного хранения и передачи информации создают условия, в которых на видеоданные приходится более 80% мирового интернет-трафика [1, p. 1]. Фиксируется экспоненциальный рост объема видеоданных [1, p. 1]. Для их обработки активно привлекаются автоматические методы, включая методы машинного обучения и искусственного интеллекта.

Во многих случаях указанные методы следуют парадигме обучения с учителем, т. е. предполагают предварительную обработку и разметку данных для обучения модели. Разметка может быть ручной, полуавтоматической или автоматической. В первом случае данные размечаются людьми вручную, во втором – при помощи алгоритмов с ручной коррекцией полученных результатов, в третьем – с помощью алгоритмов при минимальном участии человека. Ручная разметка гарантирует ясность и информативность меток, однако может требовать значительного количества человеко-часов, а также наследовать заложенные людьми субъективные искажения и ошибки. Автоматические алгоритмы менее прозрачны, но способствуют получению более быстрых и объективных результатов [2].

При обучении с учителем от содержания разметки зависит область применения обучаемых моделей. Объектом настоящей работы выступает разметка данных для решения задач в области анализа поведения, которое определяется как наблюдаемая активность людей и животных, направленная на взаимодействие с окружающей средой [3]. Работа ограничивается областью анализа поведения по видеоданным. Она сосредотачивается на действиях как компонентах поведения. Под действием понимается «простой двигательный паттерн <...>, выполнение которого занимает время <...> в пределах десятков секунд» [4]. Работа предлагает метод разметки данных для задач автоматического распознавания, классификации и прогнозирования действий по видеоданным. Предлагаемый метод относится к классу методов ручной разметки видеоданных.

## II. ОБЗОР ЛИТЕРАТУРЫ

Распознавание, классификация и прогнозирование действий по видеоданным являются востребованными задачами компьютерного зрения. Их обсуждение представляет актуальность при решении проблем в области ветеринарии, видеонаблюдения, здравоохранения, исследований потребительского поведения и др. В исследовательской литературе предложено значительное количество наборов данных и методов разметки данных для соответствующих задач [5].

Настоящая работа примыкает к исследовательской традиции, которую можно обозначить как структурный подход к определению действий. В этой традиции действие понимается как совокупность взаимосвязанных компонентов, которые обладают свойствами упорядоченности и воспроизводимости. Анализируется внутренняя организация действий, а также место действий во внутренней организации поведения. В качестве компонентов действий указываются движения, т. е. действия рассматриваются как упорядоченные и воспроизводимые комбинации движений. Признается, что одно и то же действие может быть совершено посредством сочетания разных движений. В свою очередь, действия рассматриваются как компоненты активности, или деятельности, а сменяющие друг друга формы активности – как компоненты поведения [6, p. 6-7; 7, p. 1476-1477].

Практическое использование соответствующего подхода в условиях ручной разметки видеоданных сопряжено со следующими ограничениями.

---

Исследование было проведено при поддержке Министерства науки и высшего образования РФ (соглашение № FSEE-2025-0015)

Морфологические особенности людей и животных позволяют совершать очень большое количество движений. Альтернативные варианты выполнения действий предполагают расширение алфавита движений при разметке данных. Определения движений, действий и активностей зависят от содержания решаемой задачи и могут пересматриваться. Список потенциальных задач остается открытым. Это существенно снижает шансы на получение полного и непротиворечивого перечня движений в составе действий, действий в составе активностей, а также активностей в составе поведения.

С учетом указанных ограничений предлагаемый в настоящей работе метод не содержит критериев выделения действий в видеоданных. Вместо этого он фокусируется на оценке правильности хода выполнения действий. Предлагается соотносить ход выполнения действий с алгоритмами выполнения действий, а также устанавливать степень соответствия хода выполнения действий выбранным алгоритмам. Общее направление работы выбирается с принятием во внимание актуальности проблемы «черного ящика» при использовании искусственного интеллекта и принципов объяснимого искусственного интеллекта [8].

### III. МЕТОДОЛОГИЯ И РЕЗУЛЬТАТЫ

В настоящей работе предлагается следующая формулировка цели ручной разметки данных для автоматического распознавания, классификации и прогнозирования действий по видеоданным. Даны набор обозначений действий  $C = \{c_1, \dots, c_k\}$  и набор обозначений движений  $D = \{d_1, \dots, d_m\}$ . Каждому обозначению действия  $c_i \in C$  соответствует набор алгоритмов выполнения действия  $A_i = \{a_{i,1}, \dots, a_{i,n}\}$ , где каждый алгоритм – это кортеж обозначений движений в составе действия. Для каждого алгоритма  $a_{i,n} \in A_i$  в каждом  $A_i$  сформулирована модель отклонений от алгоритма. Также дан набор видеоданных  $V = \{v_1, \dots, v_p\}$ . Каждый элемент этого набора  $v_j \in V$  состоит из последовательности кадров  $F_j = \{f_{j,1}, \dots, f_{j,q}\}$ . Кадры могут содержать информацию о ходе выполнения действий, представленных в  $C$ , включая информацию о ходе выполнения движений в составе этих действий. Для каждого  $c_i \in C$  в каждом  $v_j \in V$  необходимо выполнить следующие операции. Выделить в  $v_j$  множество интервалов кадров  $S_{ij}$ , содержащих информацию о ходе выполнения действия, обозначаемого с помощью  $c_i$ . Выделить в  $S_{ij}$  множество подынтервалов кадров  $T_{ij}$ , содержащих информацию о ходе выполнения движений в составе действия. Произвести процедуру оценки правильности выполнения действия по информации в составе  $T_{ij}$  и зафиксировать результат в формате ручной разметки.

Предлагаемый метод ручной разметки видеоданных предусматривает следующую процедуру оценки правильности выполнения действия. На первом шаге производится поиск наилучшего соответствия между информацией о ходе выполнения действия в составе  $T_{ij}$  и набором алгоритмов выполнения действия  $A_i$ . Выбирается алгоритм  $a^* \in A_i$ , для которого мера соответствия ходу выполнения действия максимальна. Если таких алгоритмов несколько, выбирается один алгоритм. На втором шаге производится анализ отклонений хода выполнения действия от алгоритма  $a^*$ .

Метод предусматривает следующие классы отклонений от алгоритма  $a^*$ .

- Пропуск необходимых движений, т. е. движений в составе алгоритма  $a^*$ .
- Наличие опциональных движений, т. е. движений, которые не являются частью алгоритма  $a^*$  и не влияют на ход выполнения действия.
- Изменение порядка выполнения движений, т. е. пропуск необходимого движения в подынтервале кадров, где оно должно быть совершено, с выполнением в другом подынтервале кадров в статусе опционального движения.
- Некорректное выполнение необходимых движений, т. е. использование такого способа выполнения необходимых движений, который противоречит алгоритму  $a^*$ .
- Наличие контрпродуктивных движений, т. е. движений, которые запрещаются алгоритмом  $a^*$  и отрицательно влияют на ход выполнения действия.

По результатам анализа отклонений хода выполнения действия от алгоритма  $a^*$  подынтервалам кадров в составе  $T_{ij}$  присваиваются метки. Если подынтервал содержит информацию о необходимом движении, которое выполняется без отклонений от алгоритма  $a^*$ , ему присваивается метка корректно выполненного движения. При наличии отклонений присваиваются метки, соответствующие характеру этих отклонений.

С целью учета человеческого фактора в работе над ручной разметкой данных предлагаемый метод предусматривает анализ содержания полученной разметки на предмет валидности и надежности. Выбор подхода к проведению соответствующего анализа выносится за рамки предлагаемого метода и признается этапом формулирования моделей отклонения от алгоритмов выполнения действий.

Общая схема практического применения предлагаемого метода представлена на рисунке 1. Разметка видеоданных начинается с определения интервалов кадров, в которых содержится информация об искомым действиях. По умолчанию кадры в границах интервалов отмечаются как содержащие опциональные движения в составе действий. Производится выбор алгоритмов для оценки правильности выполнения действий. В границах интервалов кадров с информацией о действиях выделяются подынтервалы, соответствующие необходимым шагам выбранных алгоритмов. По умолчанию им присваиваются метки о корректном выполнении. При наличии отклонений от выбранных алгоритмов метки изменяются сообразно характеру отклонений. Далее анализируются оставшиеся подынтервалы кадров. При необходимости соответствующие им метки опциональных движений корректируются. После разметки всех выделенных интервалов кадров производится проверка валидности и надежности меток. Затем работа завершается или повторяется на материале следующего набора данных.



Рис. 1. Основное содержание предлагаемого метода ручной разметки данных

Особенности использования предлагаемого в настоящей работе метода на материале конкретных данных позволяет прояснить следующий пример. Дана информация о поведении участника делового чаепития. Участник чаепития взаимодействует с чайной парой. Он поднимает чашку за ручку, отставив мизинец; подносит чашку ко рту; улыбается, глядя вверх чашки; дует на чай; смотрит в чашку и ставит ее на блюдце. Задача ручной разметки формулируется как анализ действия «прием чая». Действие соотносится с последовательностью движений, сопровождающих поднесение чайной чашки ко рту, отпитие чая из чашки и возвращение чашки на блюдце. Фиксируется наличие действия в составе анализируемых данных, а также совпадение интервала кадров в данных с интервалом кадров, содержащих информацию о действии.

Оценка правильности выполнения действия производится с опорой на нормы делового этикета [9, р. 90–108]. Формулируется следующий алгоритм выполнения действия:

- а) взять чашку за ручку, не отставляя мизинец;
- б) поднести чашку ко рту;
- в) направить взгляд в чашку;
- г) не дуть на чай;
- д) отпить чай;

е) поднять взгляд;

ж) вернуть чашку на блюдце.

Альтернативные алгоритмы выполнения действия не предусматриваются. Шаги в составе алгоритма рассматриваются как алфавит движений для разметки данных. Им присваивается буквенный код в соответствии с порядком следования движений в алгоритме. Движения, не представленные в алфавите, относятся к категории «прочие» и кодируются буквой «п». Формулируется следующая модель отклонений от алгоритма. Движения, санкционированные алгоритмом, но отсутствующие в данных, признаются пропущенными. Движения, не указанные в алгоритме, признаются опциональными. Движения, выполняемые за пределами предусмотренного алгоритмом шага, признаются выполненными со сменой порядка следования. Движения, обозначаемые буквой «а», признаются некорректными, если не соблюдены все условия выполнения. Движения, запрещаемые на шаге «г», признаются контрпродуктивными. Способы выполнения движений в составе действия кодируются числами: 0 – корректное, 1 – пропущенное, 2 – опциональное, 3 – с изменением порядка следования, 4 – некорректное, 5 – контрпродуктивное.

Производится поиск подынтервалов кадров с информацией о движениях в составе действия. Результаты сохраняются в соответствующей таблице. Строки, или записи, таблицы соотносятся с движениями. Столбцы, или поля, таблицы – с атрибутами движений. Предусматриваются следующие поля: порядковый номер, кодовое обозначение, первый кадр подынтервала, последний кадр подынтервала, способ выполнения. Выборочные результаты разметки движений в составе анализируемого действия приводятся в Таблице 1.

Способ выполнения движений в подынтервалах категории «п» по умолчанию устанавливается равным 2. В анализируемых данных к соответствующей категории относятся мимические движения участника делового чаепития при улыбке. В рамках примера разметка признается валидной и надежной. Производится завершение работы.

ТАБЛИЦА 1. ПРИМЕР ВЫБОРОЧНЫХ РЕЗУЛЬТАТОВ РУЧНОЙ РАЗМЕТКИ ДАННЫХ

№	Кодовое обозначение	Способ выполнения
1	а	4
2	б	0
3	п	2
4	в	1
5	г	5
6	д	1
7	е	1
8	в	3
9	ж	0

Приведенный вариант разметки допускает сопоставление с существующими альтернативами. В частности, с методом разметки видеоданных о чайных церемониях в составе набора данных FineTea [10]. Авторы набора данных отмечают значимость и новизну задачи по распознаванию деталей действий в составе чайных церемоний и предлагают различение 9 классов действий, организованных в 31 подкласс. Классы действий включают переворачивание чашек, подачу чая и др. Подклассы зависят от классов. Например, в

классе подачи чая представлены подклассы подачи зеленого чая, подачи чая улун и подачи черного чая. Разметка данных осуществляется по следующему принципу. Видеоданные из набора соотносятся с тремя видами активностей в составе чайных церемоний. Виды активностей различаются на основании используемого в них чая (зеленый, улун, черный). Активности подразделяются на действия. В случае зеленого чая выявляется 8 действий, в случае чая улун – 13, в случае черного чая – 10. Эти действия составляют подклассы набора данных. Затем они обобщаются в 9 базовых инвариантов, которые соответствуют классам действий в наборе данных [10].

Обращают на себя внимание следующие различия между методом разметки набора данных FineTea и методом, предлагаемым в настоящей работе. Содержание первого метода определяется разметкой действий в составе активностей; содержание второго – разметкой движений в составе действий. Первый метод использует данные о ходе выполнения действий для выявления классов действий; второй – алгоритмы выполнения действий для анализа хода выполнения действий. Первый метод решает задачу разметки данных путем объединения действий в классы; второй – путем оценки правильности выполнения действий. Перечисленные различия обуславливают разницу в области применения указанных методов.

#### IV. ЗАКЛЮЧЕНИЕ

В настоящей работе предлагается метод ручной разметки данных для применения в задачах распознавания, классификации и прогнозирования действий по видеоданным. Он примыкает к структурному подходу в области анализа поведения и предполагает принятие во внимание принципов объяснимого искусственного интеллекта. Предлагаемый метод предусматривает соотнесение хода и алгоритмов выполнения действий и включает следующие основные этапы ручной разметки видеоданных. Во-первых, поиск в видеоданных интервалов кадров, содержащих информацию о действиях. Во-вторых, выбор алгоритмов для оценки хода выполнения действий. В-третьих, поиск подынтервалов кадров с информацией о движениях в составе действий. В-четвертых, оценку степени соответствия движений в составе действий алгоритмам выполнения действий. Наконец, в-пятых, оценку валидности и надежности разметки. Использование метода позволяет извлекать из данных детальную информацию о ходе выполнения действий, а также о специфике отклонений от алгоритмов выполнения

действий. Полученную информацию можно использовать для подготовки моделей на базе машинного обучения и искусственного интеллекта, которые способны объяснять и обосновывать свои решения в области распознавания, классификации и прогнозирования действий по видеоданным. Предлагаемый метод может быть применен в прикладных задачах анализа действий по видеоданным, включая задачи в области мониторинга техники безопасности, социальной робототехники, этологии и др. Направления дальнейших исследований включают автоматизацию процедуры разметки, а также дальнейшее уточнение метода в соответствии со спецификой прикладных задач распознавания, классификации и прогнозирования действий.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Do T.-T., Huynh Q.-T., Kim K., Nguyen V.-Q. A Survey on Video Big Data Analytics: Architecture, Technologies, and Open Research Challenges // *Applied Sciences*. 2025. Vol. 15. No. 14. P. 1-40. DOI: <https://doi.org/10.3390/app15148089>.
- [2] von Ziegler L., Sturman O., Bohacek J. Big Behavior: Challenges and Opportunities in a New Era of Deep Behavior Profiling // *Neuropsychopharmacology Reviews*. 2021. Vol. 46. P. 33-44. DOI: <https://doi.org/10.1038/s41386-020-0751-7>.
- [3] Bergner R. M. What is Behavior? And so What? // *New Ideas in Psychology*. 2011. Vol. 29. P. 147-155.
- [4] Turaga P., Chellappa R., Subrahmanian V. S., Udrea O. Machine Recognition of Human Activities: A Survey // *IEEE Transactions on Circuits and Systems for Video Technology*. 2008. Vol. 18. Iss. 11. P. 1473-1488. DOI: <https://doi.org/10.1109/TCSVT.2008.2005594>.
- [5] Heo S., Moon J., Jung S. K. Action Recognition: A Comprehensive Survey of Tasks, Methods, and Challenges // *ICT Express*. 2026. Vol. 12. Iss. 1. P. 32-49. DOI: <https://doi.org/10.1016/j.ict.2025.11.015>.
- [6] Biliński P. T. Human Action Recognition in Videos. PhD Thesis to Obtain ... PhD of Science. Nice, 2014. 244 p.
- [7] Krüger V., Kragic D., Ude A., Geib Ch. The Meaning of Action: A Review on Action Recognition and Mapping // *Advanced Robotics*. 2007. Vol. 21. Iss. 13. P. 1473-1501. DOI: <https://doi.org/10.1163/156855307782148578>.
- [8] Ali S., Abuhmed T., El-Sappagh Sh., Muhammad Kh., Alonso-Moral J. M., Confalonieri R., Guidotti R., Del Ser J., Díaz-Rodríguez N., Herrera F. Explainable Artificial Intelligence (XAI): What We Know and What is Left to Attain Trustworthy Artificial Intelligence // *Information Fusion*. 2023. Vol. 99. P. 1-52. DOI: <https://doi.org/10.1016/j.inffus.2023.101805>.
- [9] Johnson D. *Tea and Etiquette: Taking Tea for Business and Pleasure* / Introduction by J. N. Pratt. Rev. ed. Sterling: Capital Books, Inc., 2002. 160 p.
- [10] Ouyang Ch., Yi Y., Wang H., Zhou J., Tian T. FineTea: A Novel Fine-Grained Action Recognition Video Dataset for Tea Ceremony Actions // *Journal of Imaging*. 2024. Vol. 10. Iss. 9. P. 1-17. DOI: <https://doi.org/10.3390/jimaging10090216>.