Юрий Алексеевич ЕГОРОВ[1]
Ирина Гелиевна ЗАХАРОВА[2]

УДК 004.93

КОНВЕЙЕРНЫЙ МЕТОД ДЛЯ РАСПОЗНАВАНИЯ КОМПЛЕКСНЫХ ДЕЙСТВИЙ ОБЪЕКТОВ В СИСТЕМАХ ВИДЕОНАБЛЮДЕНИЯ

[1] аспирант кафедры программного обеспечения,
Тюменский государственный университет
stud001651168@study.utmn.ru; ORCID: 0000-0001-7670-5283

[2] кандидат физико-математических наук,
профессор кафедры программного обеспечения,
Тюменский государственный университет
i.g.zakharova@utmn.ru

Аннотация

Разработка интеллектуальных систем видеонаблюдения — это область активных исследований, в которой представлены решения для использования в определенных условиях. Кроме того, сформулирован ряд проблем, которые требуют решения. В частности, это проблема распознавания комплексных действий, которые состоят из последовательностей элементарных действий и, как правило, трудно поддаются классификации по одному кадру видеозаписи.

Настоящее исследование посвящено решению задачи распознавания комплексных действий на видеозаписях. Целью работы является разработка конвейерного метода (пайплайна) для распознавания комплексных действий, которые совершает наблюдаемый объект на видеозаписях. Новизна работы заключается в подходе к моделированию действия с помощью последовательностей элементарных действий и сочетания нейронных сетей и стохастических моделей. Предлагаемое решение может быть использовано для разработки интеллектуальных систем видеонаблюдения с целью обеспечения безопасности на производственных объектах, включая объекты нефтегазовой отрасли.

Цитирование: Егоров Ю. А. Конвейерный метод для распознавания комплексных действий объектов в системах видеонаблюдения / Ю. А. Егоров, И. Г. Захарова // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. 2022. Том 8. № 2 (30). С. 165-182.
DOI: 10.21684/2411-7978-2022-8-2-165-182

Было проведено исследование видеозаписей объектов, совершающих различные действия. Выделены признаки, описывающие комплексные действия, и их свойства. Сформулирована задача распознавания комплексных действий, представленных последовательностью элементарных действий. В результате был разработан пайплайн, реализующий комбинированный подход. Элементарные действия описываются с помощью скелетной модели в графической форме. Каждое элементарное действие распознается с помощью сверточной нейронной сети, затем комплексные действия моделируются с помощью скрытой марковской модели. Разработанный пайплайн был протестирован на видеозаписях студентов, действия которых были разделены на две категории: списывание и обычные действия. В результате экспериментов точность классификации элементарных действий составила 0,69 по метрике accuracy, точность бинарной классификации комплексных действий составила 0,71.

Кроме того, были указаны ограничения разработанного пайплайна и выделены дальнейшие пути развития и исследования применяемых подходов, в частности исследование помехоустойчивости.

Ключевые слова

Машинное обучение, компьютерное зрение, распознавание действий, распознавание комплексных действий, сверточные нейронные сети, скрытая марковская модель, стохастические модели.

DOI: 10.21684/2411-7978-2022-8-2-165-182

Введение

Разработка систем распознавания действий является областью активных исследований. Это обусловлено широким применением разработанных решений для создания интеллектуальных систем видеонаблюдения или для организации интерфейсов взаимодействия людей с различными информационными системами и ассистентами.

Такие системы применяются для мониторинга и изучения популяций животных [28], разработки роботизированных ассистентов для детей с расстройствами аутистического спектра [30], разработки системы для облегчения мониторинга состояния пациентов в стационарах больниц [12].

Особый интерес представляет возможность использования систем распознавания действий для обеспечения безопасности на производственных объектах, таких как операционные [23] и объекты возведения сложных конструкций [38], включая объекты инфраструктуры нефтегазовой отрасли. Также такие системы могут использоваться для мониторинга действий операторов на производственных объектах с целью выявления нарушений или нештатных ситуаций.

Особенностью данной работы является то, что в ней рассматривается задача распознавания комплексных действий, состоящих из последовательности элементарных действий. Под элементарными действиями понимаются действия,

которые могут быть распознаны по одному кадру. Например, бег, письмо в тетради, езда на велосипеде. Комплексные действия — это действия, которые состоят из последовательности элементарных, и не могут быть распознаны по одному кадру. Например, списывание на экзамене, выполнение определенного элемента в фигурном катании, выполнение определенной последовательности действий при совершении каких-либо операций на рабочем месте. На рис. 1 показаны примеры, иллюстрирующие элементарное (сверху) и комплексное (снизу) действия.

Рис. 1. Сверху: пример простейшего действия «студенты пишут», которое может быть распознано по любому из представленных кадров. Снизу: пример комплексного действия «студенты начинают разговор». Особый интерес представляет момент, когда один из студентов инициирует разговор, а второй его поддерживает, т. к. его важно отличать, например, от ситуации, в которой другой студент не поддерживает призыв к началу разговора

Fig. 1. Above: an example of a simple action “students writing” that can be recognized from any of the frames presented. Below: an example of a complex action “students start a conversation”. Of particular interest is the point when one of the students initiates a conversation, and the second one supports it, since it is important to distinguish it, for example, from a situation in which another student does not support the call to start a conversation

Обзор

Конвейерные методы, или пайплайны, для решения задач распознавания действий на видеозаписях разрабатываются с использованием методов, принадлежащих к одной из четырех групп: методам машинного обучения, методам глубокого обучения, вероятностным методам, комбинированным методам.

Ввиду особенностей предметной области наиболее популярными методами, которые используют для решения задач распознавание действий, являются методы глубокого обучения.

В первую очередь, это сверточные нейронные сети (convolutional neural networks, CNN), применение которых показано в работах [12, 25]. Работа [25] примечательна тем, что исследователям с успехом удалось применить для распознавания частей тела человека сеть, предварительно обученную на синтетических данных. Также CNN применяются для разработки автоэнкодеров [6, 19]. Авторы работы [19] использовали автоэнкодеры для избавления от шумов и незначимой информации на изображениях. Исследователи в [15] разработали на базе генеративно-состязательной сети (generative adversarial network, GAN) и рекуррентной сети (recurrent neural network, RNN) комплексную модель автоэнкодера, которая позволяет описывать последовательности поз человека.

Рекуррентные сети позволяют учитывать временные признаки при распознавании действий. В [32] представлена система из двух рекуррентных сетей, одна из которых — attentional recurrent relational network-LSTM, извлекающая признаки, описывающие позу человека на отдельно взятом кадре, другая — long short-term memory (LSTM) network, распознающая действие по последовательности признаков, извлеченных первой сетью.

Естественным средством для представления поз человека является скелетная модель, которую можно рассматривать как граф. В связи с этим для решения задач распознавания действий также широко применяются графовые нейронные сети [34, 38]. Исследователями в данной области разработаны подходы, позволяющие осуществлять поиск как оптимальной топологии графа в процессе обучения сети [17], так и кадров, значимых для распознавания действий [33].

Несмотря на доминирующее положение методов глубокого обучения, методы машинного обучения и статистического анализа также успешно применяются для разработки пайплайнов. Например, в работе [35] представлены пайплайны, использующие в качестве признаков гистограммы градиентов движений (histograms of motion gradients, HMG) и алгоритмы машинного обучения для распознавания действий спортсменов. Аналогичный подход был использован в работе [9]. В исследовании [5] авторы изучали эффективность применения мешка визуальных слов (bag of visual words, BOVW) и методов машинного обучения для решения задачи классификации действий. В работе [22] авторы использовали методы статистического анализа для выбора признаков, значимых при классификации действий, и марковские модели максимальной энтропии для классификации последовательностей кадров.

Отдельный интерес вызывают пайплайны, использующие методы как глубокого, так и машинного обучения. В работе [18] предложен метод, в котором CNN обучается на признаках, представленных в виде мешка визуальных слов и векторов Фишера. В работе [37] исследователи разработали пайплайн, который использует комитет CNN и метод Байеса для агрегации результатов, полученных от различных сетей. В [36] использован комплексный подход, в котором использовалась CNN для извлечения признаков в каждом кадре и линейная динамическая система для установления временных связей между признаками, относящимися к различным кадрам.

Разнообразие подходов к решению задачи распознавания действий обусловлено большим количеством особенностей и условий, в которых решается задача.

Одной из основных особенностей является то, что одни и те же действия, за которыми наблюдают с разных ракурсов, выглядят по-разному [10]. Решение данной проблемы имеет два различных подхода. В [24] исследователи предложили описывать координаты ключевых точек скелета человека в трехмерном пространстве относительно точки обзора. Точка вычисляется для каждого кадра таким образом, чтобы скелет, описывающий действие, был инвариантным. Авторы работы [21] исследовали метод преобразования скелета так, чтобы одинаковые действия имели одинаковое описание. Отдельного рассмотрения заслуживают частные случаи, связанные с распознаванием действий от первого лица [14] и в условиях движущейся камеры [26].

При разработке пайплайнов также важно учитывать вариативность совершаемых действий. Данный вопрос рассматривали, в частности, авторы работы [27]: они разработали иерархическую динамическую модель, которая использует вероятностные методы для подбора параметров таким образом, чтобы учитывать разные вариации одних и тех же действий.

Кроме того, при распознавании действий особое место занимает вопрос поиска и описания взаимосвязей между признаками, характеризующими движение. При этом отдельно рассматриваются вопросы установления связи между пространственными и временными признаками. Например, авторы в [20] предлагают двунаправленную рекуррентную сеть для установления соответствия между временными и пространственными признаками и отсечения незначимых признаков, а также отдельно рассматривают вопросы связи между пространственными признаками в пределах одного кадра. В работе [7] исследователи предлагают представить скелет человека в кадре в виде направленного ациклического графа и использовать направленную графовую нейронную сеть, которая обучается распознавать движения с учетом взаимосвязей суставов и костей скелета. Другие авторы [31] разработали специальные структуры для локального (в рамках одного кадра) и глобального (в рамках всего видео) представления скелета и использовали данную структуру для обучения сверточной нейронной сети.

Также важно установить, какие признаки являются дискриминативными. В работе [8] разработан модуль преобразования информации о скелете человека и движении скелета между кадрами. Данный модуль при преобразовании осуществляет выбор значимых для распознавания действий суставов и возвращает информацию в формате, который позволяет проводить обучение сверточных нейронных сетей. В [29] исследователи рассматривают задачу выбора значимых признаков комплексно: поиск значимых для распознавания движений элементов скелета и поиск значимых кадров в видео. В [11] на основе генетического алгоритма разработан низкоуровневый метод отбора значимых признаков. Целью данного метода является уменьшение размерности пространства векторов, описывающих признаки, путем выбора значимых измерений.

Совершая действия, объект может взаимодействовать с другими объектами в кадре, что также является важным фактором, который необходимо учитывать при распознавании действий [13].

Выбор типа исходных данных оказывает влияние на точность распознавания как действий объектов, так и состояния объектов в широком смысле. В [37] проведено исследование информативности термограмм для распознавания психоэмоционального состояния объекта. Авторы в [4] показали, что использование изображений в инфракрасном спектре позволяет улучшить точность распознавания действий в условиях неравномерного освещения. В работе [16] авторы предлагают использовать генеративную нейронную сеть для восстановления инфракрасного изображения по цветному и, наоборот, цветного по инфракрасному.

Обзор работ в области распознавания действий показал, что в данном направлении активно проводятся исследования и сформулирован ряд проблем, требующих рассмотрения.

Методология

При разработке пайплайна был проведен анализ видеозаписей, на которых наблюдаемые объекты выполняют различные действия. В ходе анализа были выделены ключевые свойства, которыми обладают совершаемые действия:

- каждое действие может состоять из нескольких элементарных действий;
- переход от одного элементарного действия к другому имеет закономерности, которые могут быть выражены с помощью вероятностей;
- каждое элементарное действие характеризуется положением тела исполнителя и характером взаимодействия исполнителя с окружающими предметами.

В результате была сформулирована новая задача распознавания комплексных действий, состоящих из последовательностей элементарных действий.

Задача распознавания комплексных действий

Задача распознавания комплексных действий сформулирована как задача классификации последовательностей элементарных действий.

Дано множество $X = \{x: x$ — описание элементарного действия$\}$; конечное множество $Y = \{y: y$ — категория элементарного действия$\}$. При этом категория элементарного действия $y_i \in Y$ соответствует описанию $x_j \in X$. Множество $Z = \{z: z$ — комплексное действие$\}$, каждое комплексное действие $z_i \in Z$ характеризуется последовательностью элементарных действий $z_i = \{y_i^j\}^K$, где K — количество элементарных действий в последовательности, определяющей комплексное действие, задано априорно.

Необходимо построить классификатор $F_1: X \to Y$, который каждому описанию элементарного действия $x \in X$ сопоставляет верную категорию элементарного действия $y \in Y$, построить классификатор $F_2: Y^K \to Z$, который последовательности элементарных действий $y_i = \{y_i^j\}^K$ сопоставляет комплексное действие объекта $z \in Z$.

В работе рассматривается применение комбинированного подхода. Предлагаемый подход использует CNN и вероятностные методы для распознавания последовательностей действий. В том числе рассматриваются проблемы выбора и представления признаков для распознавания действий с помощью CNN.

Результаты

Для решения сформулированной задачи был разработан пайплайн, состоящий из трех этапов: предобработка кадров видеопотока, распознавание элементарных действий, распознавание комплексных действий.

Предобработка кадров

На первом этапе происходит выделение признаков, описывающих элементарные действия.

Элементарное действие описывается скелетной моделью. Скелетная модель объекта $x \in X$ представляет собой граф G_x, состоящий из множества вершин V_x и множества ребер E_x. Множество вершин определяется как $V_x = \{v_x: v_x$ — координаты опорной точки$\}$, множество ребер определяется как $E_x = \{(v_x, u_x): v_x, u_x \in V_x\}$.

Для распознавания используется графическая форма скелетной модели, нанесенная на исходное изображение. Рис. 2 иллюстрирует пример изображения, прошедшего предобработку описанным методом. В работе [2] экспериментально было показано, что использование данного метода предобработки позволяет повысить точность распознавания элементарных действий с помощью сверточных нейронных сетей.

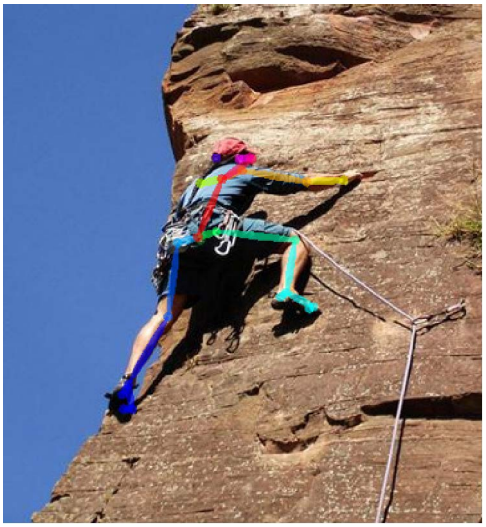

Рис. 2. Пример кадра с нанесенной скелетной моделью в графической форме

Fig. 2. An example of the frame with the drawn skeleton model in graphical representation

Распознавание элементарных действий

Распознавание промежуточных состояний осуществляется с помощью классификатора $F_1: X \to Y$, который представляет собой сверточную нейронную сеть. Предложенная сеть включает в себя блоки двух типов: блок сверточных слоев и блок полносвязных слоев.

Блок сверточных слоев состоит из последовательности слоев $C — BN — C — BN — P$, где C — сверточный слой, BN — слой нормализации, P — слой пулинга. После каждого сверточного слоя применяется функция активации LeakyReLU.

Полносвязный слой состоит из последовательности слоев $FL_1 — BN — FL_2$, где FL_1 и FL_2 — полносвязные слои. В качестве функции активации используется ReLU.

В таблице 1 представлены основные параметры блоков, составляющих нейронную сеть. Conv_1, Conv_2 — блоки сверточных слоев; Fk_1, Fk_2, Fk_3 — блоки полносвязных слоев.

Для обучения сети использовался алгоритм Adam, обучение осуществлялось в течение 30 эпох с убывающей скоростью обучения в диапазоне от (10^{-2}, 10^{-4}).

Таблица 1

Структура сверточной сети для классификации элементарных действий

Table 1

Structure of convolutional neural network for simple actions classification

Тип блока	Количество блоков	Количество сверточных слоев	Ядро свертки	Ядро пулинга	Количество нейронов
Conv_1	3	64	5×5	2×2	—
Conv_2	3	128	3×3	2×2	—
Fk_1	1	—	—	—	50
Fk_2	1	—	—	—	10
Fk_3	1	—	—	—	5

Классификация комплексных действий

Комплексное действие каждого типа моделируется и распознается с помощью скрытой марковской модели (СММ) $F_2(z)$, где $z \in Z$ — категория комплексного действия. Каждая СММ задается параметрами (N, R, P, P_0, F_1), где:

- N — множество скрытых состояний СММ;
- R — распределение вероятности того, что объект совершает действие категории $z \in Z$, моделируемое заданной СММ, находясь в состоянии $n^j \in N$;
- P — распределение вероятности перехода в состояние n^{j+1} из текущего состояния n^j;

— P_0 — распределение вероятностей начального состояния n^0;

— $F_1: X \to Y$ — классификатор, который каждому описанию элементарного действия $x \in X$ сопоставляет категорию элементарного действия $y \in Y$.

Каждая СММ, моделирующая комплексное действие определенной категории, действует по заданному итерационному алгоритму.

1. Создается классификатор элементарных действий (промежуточных состояний) F_1.
2. На шаге $j = 0$ создается классификатор $F_2(z)$ с параметрами: F_1 — классификатор, $n^0 \sim P_0$ — начальное состояние.
3. На шаге $j = \overline{(1, K-1)}$ 1:
   — по описанию x^j классификатор F_1 находит категорию состояния $y^j \in Y$;
   — СММ возвращает вероятность $p \sim P(z \mid n^{j-1}, n^j)$ — степень уверенности в том, что обрабатываемая последовательность элементарных действий принадлежит к категории z;
   — СММ переходит в состояние y^j.

Каждая СММ $F_2(z)$ возвращает вероятность принадлежности комплексного действия к категории z. Конечный результат рассчитывается по максимальной вероятности принадлежности действия некоторой категории $\underset{z}{\operatorname{argmax}}\{p(z): z \in Z\}$.

Результаты экспериментов

Для оценки разработанного пайплайна была проведена серия экспериментов, целью которых являлось нахождение точности классификации комплексных действий.

Материалом для экспериментов послужили видеозаписи, на которых были запечатлены действия студентов на занятиях. Было использовано 70 видеороликов. Элементарные действия, которые выполняли студенты, были разбиты на пять классов: пишет, читает с доски, работает за компьютером, пытается привлечь внимание других студентов, разговаривает на занятии.

При этом комплексные действия были разделены на два класса: допустимые на занятии действия; действия, которые на занятии делать нельзя.

При обучении классификаторов в пайплайне использовалась схема train-validation-test. Точность оценивалась с помощью метрики accuracy.

Точность распознавания элементарных действий составила 0,69. Точность распознавания комплексных действий составила 0,71.

Обсуждение

Разработанный пайплайн имеет следующие ограничения:

— категории элементарных и комплексных действий должны быть известны заранее;

— для описания различных действий достаточно использовать одинаковый и фиксированный шаг дискретизации действий;

— в каждый момент времени имеется полное описание, достаточное для распознавания состояния объекта; достаточность описания определяется на основе экспертной оценки.

Среди перечисленных ограничений особый интерес представляет вопрос выбора метода дискретизации видео, что дает возможности для улучшения представленного решения.

Одним из недостатков реализованного подхода является необходимость строить отдельную марковскую модель для каждой категории комплексных действий. При этом на этапе классификации также необходимо обрабатывать входную последовательность каждой из моделей. Существует гипотеза о том, что с помощью операций над изоморфными графами [1] можно объединять несколько марковских моделей для распознавания сразу нескольких категорий комплексных действий. Кроме того, проведения дополнительных исследований требуют вопросы устойчивости пайплайна к помехам. В частности, к ошибкам выделения признаков элементарных действий и ошибкам классификации элементарных действий и их влиянию на точность классификации комплексных действий.

Заключение

В результате исследования был разработан пайплайн для распознавания комплексных действий, который может быть использован для создания интеллектуальных систем видеонаблюдения на производственных объектах.

Реализованный комбинированный метод обладает рядом преимуществ:

— гибкостью, что позволяет при необходимости заменять компоненты пайплайна;

— расширяемостью: новые категории комплексных действий можно добавлять без изменения всего пайплайна в целом при условии, что множество элементарных действий не изменяется;

— потенциальной устойчивостью к помехам, которая требует дополнительных исследований.

В то же время нельзя сказать, что модели, реализованные в пайплайне, являются универсальными, и их использование для конкретных случаев требует предварительных исследований. Численные эксперименты также показали необходимость доработки представленного решения.

СПИСОК ЛИТЕРАТУРЫ

1. Егоров Ю. А. Алгоритм FDET для построения пространства признаков классификации сложных объектов в рамках графовой модели / Ю. А. Егоров, М. С. Воробьёва, А. М. Воробьёв // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. 2017. Том 3. № 3. С. 125-134. DOI: 10.21684/2411-7978-2017-3-3-125-134
2. Егоров Ю. А. Стохастический метод распознавания действий человека на базе скелетной модели / Ю. А. Егоров, И. Г. Захарова, А. Р. Гасанов, А. А. Филицин // Информационные системы и технологии: тр. Восьмой Междурнар. науч. конф. 2020. С. 96-102.

3. Albanie S. BSL-1K: Scaling up co-articulated sign language recognition using mouthing cues / S. Albanie, G. Varlo, L. Momeni, T. Afouras, J. S. Chung, N. Fox, A. Zisserman // ECCV 2020: Computer Vision — ECCV 2020. 2020. Pp. 35-53. DOI: 10.48550/arXiv.2007.12131
4. Ali S. Variational learning of beta-liouville hidden Markov models for infrared action recognition / S. Ali, N. Bouguila // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). DOI: 10.1109/CVPRW.2019.00119
5. Aslan M. F. Human action recognition with bag of visual words using different machine learning methods and hyperparameter optimization / M. F. Aslan, A. Durdu, K. Sabanci // Neural Computing and Applications. 2020. No. 32. Pp. 8585-8597. DOI: 10.1007/s00521-019-04365-9
6. Bilal M. A transfer learning-based efficient spatiotemporal human action recognition framework for long and overlapping action classes / M. Bilal, M. Maqsood, S. Yasmin, N. U. Hasan, Seungmin Rho // The Journal of Supercomputing. 2022. Vol. 78. No. 2. Pp. 2873-2908. DOI: 10.1007/s11227-021-03957-4
7. Chao Li. Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation / Chao Li, Qiaoyong Zhong, Di Xie, Shiliang Pu // IJCAI'18: Proceedings of the 27th International Joint Conference on Artificial Intelligence. 2018. Pp. 786-792. DOI: 10.48550/arXiv.1804.06055
8. Chao Li. Skeleton-based action recognition with convolutional neural networks / Chao Li, Qiaoyong Zhong, Di Xie, Shiliang Pu // 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). 2017. Pp. 597-600. DOI: 10.48550/arXiv.1704.07595
9. Duta I. C. Efficient human action recognition using histograms of motion gradients and VLAD with descriptor shape information / I. C. Duta, J. R. R. Uijlings, B. Ionescu, K. Aizawa, A. G. Hauptmann, N. Sebe // Multimedia Tools and Applications. 2017. Vol. 76. No. 21. Pp. 22445-22472. DOI: 10.1007/s11042-017-4795-6
10. Ghojogh B. Fisherposes for human action recognition using kinect sensor data / B. Ghojogh, H. Mohammadzade, M. Mokari // EEE Sensors Journal. 2018. Vol. 18. No. 4. Pp. 1612-1627. DOI: 10.1109/JSEN.2017.2784425
11. Guha R. CGA: A new feature selection model for visual human action recognition / R. Guha, A. H. Khan, P. K. Singh, R. Sarkar, D. Bhattacharjee // Neural Computing and Applications. 2021. No. 33. Pp. 5267-5286. DOI: 10.1007/s00521-020-05297-5
12. Gul M. A. Patient monitoring by abnormal human activity recognition based on CNN architecture / M. A. Gul, M. H. Yousaf, S. Nawaz, Z. U. Rehman, H. Kim // Electronics. 2020. Vol. 9. No. 12. Pp. 1-14. DOI: 10.3390/electronics9121993
13. Hongsong Wang. Learning content and style: Joint action recognition and person identification from human skeletons / Hongsong Wang, Liang Wang // Pattern Recognition. Vol. 81. 2018. Pp. 23-25. DOI: 10.1016/j.patcog.2018.03.030
14. Kapidis G. Egocentric hand track and object-based human action recognition / G. Kapidis, R. Poppe, E. van Dam, L. P. J. J. Noldus, R. Veltkamp // 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). 2019. Pp. 922-929. DOI: 10.48550/arXiv.1905.00742
15. Kundu J. N. Unsupervised feature learning of human actions as trajectories in pose embedding manifold / J. N. Kundu, M. Gor, P. K. Uppala, R. V. Babu // 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). 2019. Pp. 1459-1467. DOI: 10.48550/arXiv.1812.02592

16. Lan Wang. PM-GANs: Discriminative representation learning for action recognition using partial-modalities / Lan Wang, Chenqiang Gao, Luyu Yang, Yue Zhao, Wangmeng Zuo, Deyu Meng // Proceedings of the European Conference on Computer Vision (ECCV). 2018. Pp. 384-401. DOI: 10.48550/arXiv.1804.06248
17. Lei Shi. Two-stream adaptive graph convolutional networks for skeleton-based action recognition / Lei Shi, Yifan Zhang, Jian Cheng, Hanqing Lu // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. Pp. 12026-12035. DOI: 10.48550/arXiv.1805.07694
18. Lei Wang. Hallucinating IDT descriptors and I3D optical flow features for action recognition with CNNs / Lei Wang, P. Koniusz, Du Q. Huynh // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019. Pp. 8698-8708. DOI: 10.48550/arXiv.1906.05910
19. Ludl D. Simple yet efficient real-time pose-based action recognition / D. Ludl, T. Gulde, C. Curio // IEEE Intelligent Transportation Systems Conference (ITSC). 2019. Pp. 581-588. DOI: 10.48550/arXiv.1904.09140
20. Maosen Li. Actional-structural graph convolutional networks for skeleton-based action recognition / Maosen Li, Siheng Chen, Xu Chen, Ya Zhang, Yanfeng Wang, Qi Tian // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. Pp. 3595-3603. DOI: 10.48550/arXiv.1904.12659
21. Mengyuan Liu. Enhanced skeleton visualization for view invariant human action recognition / Mengyuan Liu, Hong Liu, Chen Chen // Pattern Recognition. 2017. Vol. 68. Pp. 346-362. DOI: 10.1016/j.patcog.2017.02.030
22. Nadeem A. Accurate physical activity recognition using multidimensional features and Markov model for smart health fitness / A. Nadeem, A. Jalal, K. Kim // Symmetry. 2020. Vol. 12. No. 11. Pp. 1766-1783. DOI: 10.3390/sym12111766
23. Padoy N. Machine and deep learning for workflow recognition during surgery / N. Padoy // Minimally Invasive Therapy & Allied Technologies. 2019. Vol. 28. No. 2. Pp. 82-90. DOI: 10.1080/13645706.2019.1584116
24. Pengfei Zhang. View adaptive recurrent neural networks for high performance human action recognition from skeleton data / Pengfei Zhang, Cuiling Lan, Junliang Xing, Wenjun Zeng, Jianru Xue, Nanning Zheng // Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017. Pp. 2117-2126. DOI: 10.48550/arXiv.1703.08274
25. Rahmani H. Learning action recognition model from depth and skeleton videos / H. Rahmani, M. Bennamoun // Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017. Pp. 5832-5841. DOI: 10.1109/ICCV.2017.621
26. Rezazadegan F. Action recognition: From static datasets to moving robots / F. Rezazadegan, S. Shirazi, B. Upcrofit, M. Milford // 2017 IEEE International Conference on Robotics and Automation (ICRA). 2018. Pp. 3185-3191. DOI: 10.48550/arXiv.1701.04925
27. Rui Zhao. Bayesian hierarchical dynamic model for human action recognition / Rui Zhao, Wanru Xu, Hui Su, Qiang Ji // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. Pp. 7733-7742. DOI: 10.1109/CVPR.2019.00792
28. Schofield D. Chimpanzee face recognition from videos in the wild using deep learning / D. Schofield, A. Nagrani, A. Zisserman, M. Hayashi, M. Matsuzawa, D. Biro, S. Carvalho // Science Advances. 2019. Vol. 5. No. 9. Pp. 1-9. DOI: 10.1126/sciadv.aaw0736

29. Sijie Song. An end-to-end spatio-temporal attention model for human action recognition from skeleton data / Sijie Song, Cuiling Lan, Junliang Xing, Wenjun Zeng, Jiaying Liu // Proceedings of the AAAI Conference on Artificial Intelligence. 2017. Vol. 31. No. 1. DOI: 10.48550/arXiv.1611.06067
30. Silva V. Skeleton driven action recognition using an image-based spatial-temporal representation and convolution neural network / V. Silva, F. Soares, C. P. Leão, J. S. Esteves, G. Vercelli // Sensors. 2021. Vol. 21. No. 13. Paper 4342. DOI: 10.3390/s21134342
31. Weizhi Nie. SRNet: Structured relevance feature learning network from skeleton data for human action recognition / Weizhi Nie, Wei Wang, Xiangdong Huang // EEE Access. 2017. Vol. 7. Pp. 132161-132172. DOI: 10.1109/ACCESS.2019.2940281
32. Wu Zheng. Relational network for skeleton-based action recognition / Wu Zheng, Lin Li, Zhaoxiang Zhang, Yan Huang, Liang Wang // IEEE International Conference on Multimedia and Expo (ICME). 2019. Pp. 826-831. DOI: 10.48550/arXiv.1805.02556
33. Yansong Tang. Deep progressive reinforcement learning for skeleton-based action recognition / Yansong Tang, Yi Tian, Jiwen Lu, Peiyang Li, Jie Zhou // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. Pp. 5323-5332. DOI: 10.1109/CVPR.2018.00558
34. Yi-Fan Song. Stronger, faster and more explainable: A graph convolutional baseline for skeleton-based action recognition / Yi-Fan Song, Zhang Zhang, Caifeng Shan, Liang Wang // Proceedings of the 28th ACM International Conference on Multimedia. 2020. Pp. 1625-1633. DOI: 10.1145/3394171.3413802
35. Zhiguo Pan. Robust basketball sports recognition by leveraging motion block estimation / Zhiguo Pan, Chao Li // Signal Processing: Image Communication. 2020. Vol. 83. Paper 115784. DOI: 10.1016/j.image.2020.115784
36. Zhouning Du. Action recognition based on linear dynamical systems with deep features in videos / Zhouning Du, Hiroaki Mukaidani, Ramasamy Saravanakumar // 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2020. Pp. 2634-2639. DOI: 10.1109/SMC42975.2020.9283429
37. Zhumazhanova S. S. Statistical approach for subject's state identification by face and neck thermograms with small training sample / S. S. Zhumazhanova, A. E. Sulavko, D. B. Ponomarev, V. A. Pasenchuk // IFAC-PapersOnLine. 2019. Vol. 52. No. 25. Pp. 46-51. DOI: 10.1016/j.ifacol.2019.12.444
38. Zi-Hao Lin. Temporal image analytics for abnormal construction activity identification / Zi-Hao Lin, Albert Y. Chen, Shang-Hsien Hsieh // Automation in Construction. 2021. Vol. 124. Paper 103572. DOI: 10.1016/j.autcon.2021.103572

Yury A. EGOROV[1]
Irina G. ZAKHAROVA[2]

UDC 004.93

PIPELINE FOR COMPLEX ACTIONS RECOGNITION IN VIDEO SURVEILLANCE SYSTEMS

[1] Postgraduate Student,
Department of Software,
University of Tyumen
stud001651168@study.utmn.ru; ORCID: 0000-0001-7670-5283

[2] Cand. Sci. (Phys.-Math.), Professor,
Department of Software,
University of Tyumen
i.g.zakharova@utmn.ru

Abstract

The development of intelligent video surveillance systems is an area of active research, presenting solutions for use in specific environments. In addition, several problems have been formulated that need to be addressed. This is the problem of recognizing complex actions, which consist of sequences of elementary actions and, as a rule, are difficult to classify from a single frame of a video recording.

The present study is devoted to solving the problem of recognizing complex actions on video recordings. The aim of the work is to develop a pipeline for recognizing complex actions that an observed object performs on video recordings. The novelty of the work lies in the approach to action modeling using sequences of elementary actions and a combination of neural networks and stochastic models. The proposed solution can be used to develop intelligent video surveillance systems to ensure security at production facilities, including oil and gas industry facilities.

We analyzed video recordings of objects performing various actions. The features describing complex actions and their properties are singled out. The problem of recognition of complex

Citation: Egorov Yu. A., Zakharova I. G. 2022. “Pipeline for complex actions recognition in video surveillance systems”. Tyumen State University Herald. Physical and Mathematical Modeling. Oil, Gas, Energy, vol. 8, no. 2 (30), pp. 165-182.
DOI: 10.21684/2411-7978-2022-8-2-165-182

actions represented by a sequence of elementary actions is formulated. As a result, we developed a pipeline implements a combined approach. Elementary actions are described using a skeletal model in graphical form. Each elementary action is recognized using a convolutional neural network, then complex actions are modeled using a hidden Markov model. The developed pipeline was tested on videos of students, whose actions were divided into two categories: cheating and ordinary actions. As a result of the experiments, the classification accuracy of elementary actions was 0.69 according to the accuracy metric, the accuracy of the binary classification of complex actions was 0.71.

In addition, the constraints of the developed pipeline were indicated and further ways of enhancing the applied approaches were highlighted, in particular, the study of noise immunity.

Keywords

Machine learning, computer vision, actions recognition, complex actions recognition, convolutional neural networks, hidden Markov model, stochastic models.

DOI: 10.21684/2411-7978-2022-8-2-165-182

REFERENCES

1. Egorov Yu. A., Vorobyova M. S., Vorobyov A. M. 2017. "FDET algorithm for building space of classification patterns in graph model". Tyumen State University Herald. Physical and Mathematical Modeling. Oil, Gas, Energy, vol. 3, no. 3, pp. 125-134. DOI: 10.21684/2411-7978-2017-3-3-125-134 [In Russian]
2. Egorov Y. A., Zakharova I. G., Gasanov A. R., Filitsin A. A. 2020. "Stichastic modeling for skeleton based human action diagnostics". Information systems and technologies: Proceedings of the 8th International Scientific Conference, pp. 96-102. [In Russian]
3. Albanie S., Varlo G., Momeni L., Afouras T., Chung J. S., Fox N., A. Zisserman A. 2020 "BSL-1K: Scaling up co-articulated sign language recognition using mouthing cues". ECCV 2020: Computer Vision — ECCV 2020, pp. 35-53. DOI: 10.48550/arXiv.2007.12131
4. Ali S., Bouguila N. 2019. "Variational learning of beta-liouville hidden Markov models for infrared action recognition". Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPRW). DOI: 10.1109/CVPRW.2019.00119
5. Aslan M. F., Durdu A., Sabanci K. 2020. "Human action recognition with bag of visual words using different machine learning methods and hyperparameter optimization". Neural Computing and Applications, no. 32, pp. 8585-8597. DOI: 10.1007/s00521-019-04365-9
6. Bilal M., Maqsood M., Yasmin S., Hasan N. U., Seungmin Rho. 2020. "A transfer learning-based efficient spatiotemporal human action recognition framework for long and overlapping action classes". The Journal of Supercomputing, vol. 78, no. 2, pp. 2873-2908. DOI: 10.1007/s11227-021-03957-4
7. Chao Li, Qiaoyong Zhong, Di Xie, Shiliang Pu. 2018. "Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation". IJCAI'18: Proceedings of the 27th International Joint Conference on Artificial Intelligence, pp. 786-792. DOI: 10.48550/arXiv.1804.06055

8. Chao Li, Qiaoyong Zhong, Di Xie, Shiliang Pu. 2017. “Skeleton-based action recognition with convolutional neural networks”. 2017 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 597-600. DOI: 10.48550/arXiv.1704.07595
9. Duta I. C., Uijlings J. R. R., Ionescu B., Aizawa K., Hauptmann A. G., Sebe N. 2017. “Efficient human action recognition using histograms of motion gradients and VLAD with descriptor shape information”. Multimedia Tools and Applications, vol. 76, no. 21, pp 22445-22472. DOI: 10.1007/s11042-017-4795-6
10. Ghojogh B, Mohammadzade H., Mokari M. 2018. “Fisherposes for Human Action Recognition Using Kinect Sensor Data”. EEE Sensors Journal, vol. 18, no. 4, pp. 1612-1627. DOI: 10.1109/JSEN.2017.2784425
11. Guha R., Khan A. H., Singh P. K., Sarkar R., Bhattacharjee D. 2021. “CGA: a new feature selection model for visual human action recognition”. Neural Computing and Applications, no. 33, pp. 5267-5286. DOI: 10.1007/s00521-020-05297-5
12. Gul M. A., Yousaf M. H., Nawaz S., Rehman Z. U., Kim H. 2020. “Patient monitoring by abnormal human activity recognition based on CNN architecture”. Electronics, vol. 9, no. 12, pp 1-14. DOI: 10.3390/electronics9121993
13. Hongsong Wang, Liang Wang. 2018. “Learning content and style: Joint action recognition and person identification from human skeletons”. Pattern Recognition, vol. 81, pp. 23-25. DOI: 10.1016/j.patcog.2018.03.030
14. Kapidis G., Poppe R., van Dam E., Noldus L. P. J. J., Veltkamp R. 2019. “Egocentric hand track and object-based human action recognition”. 2019 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), pp. 922-929. DOI: 10.48550/arXiv.1905.00742
15. Kundu J. N., Gor M., Uppala P. K., Babu R. V. 2019. “Unsupervised feature learning of human actions as trajectories in pose embedding manifold”. IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1459-1467. DOI: 10.48550/arXiv.1812.02592
16. Lan Wang, Chenqiang Gao, Luyu Yang, Yue Zhao, Wangmeng Zuo, Deyu Meng. 2018. “PM-GANs: Discriminative representation learning for action recognition using partial-modalities”. Proceedings of the European Conference on Computer Vision (ECCV), pp. 384-401. DOI: 10.48550/arXiv.1804.06248
17. Lei Shi, Yifan Zhang, Jian Cheng, Hanqing Lu. 2019. “Two-stream adaptive graph convolutional networks for skeleton-based action recognition”. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12026-12035. DOI: 10.48550/arXiv.1805.07694
18. Lei Wang, Koniusz P., Huynh Du Q. 2019. “Hallucinating IDT descriptors and I3D optical flow features for action recognition with CNNs”. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Pp. 8698-8708. DOI: 10.48550/arXiv.1906.05910
19. Ludl D., Gulde T., Curio C. “Simple yet efficient real-time pose-based action recognition”. IEEE Intelligent Transportation Systems Conference (ITSC), pp. 581-588. DOI: 10.48550/arXiv.1904.09140

20. Maosen Li, Siheng Chen, Xu Chen, Ya Zhang, Yanfeng Wang, Qi Tian. 2019. "Actional-structural graph convolutional networks for skeleton-based action recognition". Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3595-3603. DOI: 10.48550/arXiv.1904.12659
21. Mengyuan Liu, Hong Liu, Chen Chen. 2017. "Enhanced skeleton visualization for view invariant human action recognition". Pattern Recognition, vol. 68, pp. 346-362. DOI: 10.1016/j.patcog.2017.02.030
22. Nadeem A., Jalal A., Kim K. 2020. "Accurate physical activity recognition using multidimensional features and Markov model for smart health fitness". Symmetry, vol. 12, no. 11, pp. 1766-1783. DOI: 10.3390/sym12111766
23. Padoy N. 2019. "Machine and deep learning for workflow recognition during surgery". Minimally Invasive Therapy and Allied Technologies, no. 28, pp. 82-90. DOI: 10.1080/13645706.2019.1584116
24. Pengfei Zhang, Cuiling Lan, Junliang Xing, Wenjun Zeng, Jianru Xue, Nanning Zheng. 2017. "View adaptive recurrent neural networks for high performance human action recognition from skeleton data". Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2117-2126. DOI: 10.48550/arXiv.1703.08274
25. Rahmani H., Bennamoun M. 2017. "Learning action recognition model from depth and skeleton videos". Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 5832-5841. DOI: 10.1109/ICCV.2017.621
26. Rezazadegan F., Shirazi S., Upcrofit B., Milford M. 2018. "Action recognition: From static datasets to moving robots". 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 3185-3191. DOI: 10.48550/arXiv.1701.04925
27. Rui Zhao, Wanru Xu, Hui Su, Qiang Ji. 2019. "Bayesian hierarchical dynamic model for human action recognition". Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7733-7742. DOI: 10.1109/CVPR.2019.00792
28. Schofield D., Nagrani A., Zisserman A., Hayashi M., Matsuzawa M., Biro D., Carvalho S. 2019. "Chimpanzee face recognition from videos in the wild using deep learning". Science Advances, vol. 5, no. 9, pp. 1-9. DOI: 10.1126/sciadv.aaw0736
29. Sijie Song, Cuiling Lan, Junliang Xing, Wenjun Zeng, Jiaying Liu. 2017. "An end-to-end spatio-temporal attention model for human action recognition from skeleton data". Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, no. 1. DOI: 10.48550/arXiv.1611.06067
30. Silva V., Soares F., Leão C. P., Esteves J. S., Vercelli G. 2021. "Skeleton driven action recognition using an image-based spatial-temporal representation and convolution neural network". Sensors, vol. 21, no. 13, paper 4342. DOI: 10.3390/s21134342
31. Weizhi Nie, Wei Wang, Xiangdong Huang. 2017. "SRNet: Structured relevance feature learning network from skeleton data for human action recognition". EEE Access, vol. 7, pp. 132161-132172. DOI: 10.1109/ACCESS.2019.2940281
32. Wu Zheng, Lin Li, Zhaoxiang Zhang, Yan Huang, Liang Wang. 2019. "Relational network for skeleton-based action recognition". IEEE International Conference on Multimedia and Expo (ICME), pp. 826-831. DOI: 10.48550/arXiv.1805.02556
33. Yansong Tang, Yi Tian, Jiwen Lu, Peiyang Li, Jie Zhou. 2018. "Deep progressive reinforcement learning for skeleton-based action recognition". Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5323-5332. DOI: 10.1109/CVPR.2018.00558

34. Yi-Fan Song, Zhang Zhang, Caifeng Shan, Liang Wang. 2020. “Stronger, faster and more explainable: A graph convolutional baseline for skeleton-based action recognition”. Proceedings of the 28th ACM International Conference on Multimedia, pp. 1625-1633. DOI: 10.1145/3394171.3413802
35. Zhiguo Pan, Chao Li. 2020. “Robust basketball sports recognition by leveraging motion block estimation”. Signal Processing: Image Communication, vol. 83. paper 115784. DOI: 10.1016/j.image.2020.115784
36. Zhouning Du, Hiroaki Mukaidani, Ramasamy Saravanakumar. 2020. “Action recognition based on linear dynamical systems with deep features in videos”. 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2634-2639. DOI: 10.1109/SMC42975.2020.9283429
37. Zhumazhanova S. S., Sulavko A. E., Ponomarev D. B., Pasenchuk V. A. 2019. “Statistical approach for subject’s state identification by face and neck thermograms with small training sample”. IFAC-PapersOnLine, vol. 52, no. 25, pp. 46-51. DOI: 10.1016/j.ifacol.2019.12.444
38. Zi-Hao Lin, Albert Y. Chen, Shang-Hsien Hsieh. 2021. “Temporal image analytics for abnormal construction activity identification”. Automation in Construction, vol. 124. paper 103572. DOI: 10.1016/j.autcon.2021.103572