

Ю.А. Егоров

Тюменский государственный университет, г. Тюмень

УДК 004.85

РАСПОЗНАВАНИЕ СОСТОЯНИЙ ДИНАМИЧЕСКИХ ОБЪЕКТОВ

Аннотация. В статье приводится описание и постановка задачи классификации последовательностей состояний динамических объектов. Также представлено описание исследуемых данных и особенности решаемой задачи. Предложена структура для представления наблюдаемых объектов.

Ключевые слова: распознавание действий, сверточные нейронные сети, компьютерное зрение, классификация последовательностей состояний.

Введение

Целью работы является предварительный анализ данных, необходимый для решения более широкой задачи – задачи классификации последовательностей состояний динамических объектов. В ходе решения задачи необходимо осуществить анализ видеозаписей, полученных в процессе наблюдения за некоторым объектом, который может изменять свое состояние с течением времени, и разработать модели и алгоритмы, позволяющие в автоматизированном режиме проводить классификацию этих состояний.

Решение данной задачи может найти широкое применение в таких сферах как:

- классификация наблюдаемых объектов по характерным для них последовательностям состояний;
- сфера образования (обучение какому-либо виду деятельности, связанному с выполнением движений);

- поиск и систематизация информации (классификация видео, расстановка тегов, разбиение видео на части, поиск ключевых эпизодов в видео);
- выявление нештатных ситуаций во время проведения мероприятий и в местах большого скопления людей.

Предварительный обзор исследований по данной тематике показал, что активная работа над решением задач распознавания действий ведется с начала 2000-х годов, но существенные результаты в данной области были получены сравнительно недавно с развитием нейронных сетей и появлением таких архитектур нейронных сетей как сверточные сети, пространственные пирамидальные сети [1], двухпоточковые сети и мультизадачное обучение [2] и др. В то же время такие проблемы, как классификация последовательностей состояний динамических объектов, изучены не в полной мере. Также особое значение в работе имеет специфика используемых данных и их подготовка для анализа.

Постановка задачи

Прежде чем давать формальную постановку, приведем описание прикладной задачи. Даны видеозаписи, на которых запечатлен процесс сдачи учениками выпускных экзаменов. На данных видеозаписях необходимо распознать последовательности состояний каждого участника экзамена (ученики, организаторы, наблюдатели и др.). Данные состояния выражаются действиями участников экзамена. Среди этих состояний некоторого участника экзамена необходимо выделить последовательность состояний, которую можно классифицировать как действия, нарушающие регламент проведения экзамена, либо как действия, соответствующие тому, что во время проведения экзамена возникла нештатная ситуация.

Таким образом, осуществляется наблюдение за некоторым множеством объектов $Z = \{z_1, z_2, \dots, z_n\}$, где $|Z| = n$. Каждый объект z_i характеризуется последовательностью состояний $X_i = \{x(t)_i^j\}_{j=1}^m$, где t – некоторый момент

времени, $i \in [1, n]$ – номер соответствующего объекта z_i , $j \in [1, m]$ – номер элемента последовательности.

Необходимо найти отображение $\phi: X \rightarrow X'$, которое из каждой последовательности X_i выделяет подпоследовательность X'_i , по некоторому правилу, и отображение $\psi: X' \rightarrow Y$, сопоставляющее подпоследовательности X'_i , метку класса $y \in Y = \{y_1, y_2, \dots, y_l\}$, где Y – множество непересекающихся классов, $|Y| = l$.

Анализ исследуемых данных

Был проведен анализ исследуемых данных, в ходе которого были выявлены следующие особенности.

Мимику участников экзамена использовать для распознавания состояний затруднительно из-за низкого качества видеозаписей, поэтому в первую очередь, при распознавании состояний необходимо основываться на движениях тела и перемещениях участников экзамена. Каждый участник экзамена снимается двумя камерами, неподвижно закрепленными на определенных местах, что исключает необходимость учитывать движение камеры при распознавании действий (Рис. 1). При этом необходимо учитывать то, что участники экзамена снимаются в разных ракурсах, кроме того, ракурс может меняться в ходе съемки в зависимости от перемещений участника по аудитории. Во время проведения экзамена необходимо наблюдать за конечным множеством участников, которое определяется до его начала, другие лица попадают в объективы камеры в исключительных ситуациях. При распознавании состояний участников экзамена важно учитывать класс участника – ученик, организатор, наблюдатель – так как для участника определенного класса существует характерное для него множество допустимых и недопустимых состояний (движения, перемещения по аудитории, время пребывания в кадре). Также важно учитывать то, что участники экзамена могут взаимодействовать как между собой, так и с окружающими их предметами.

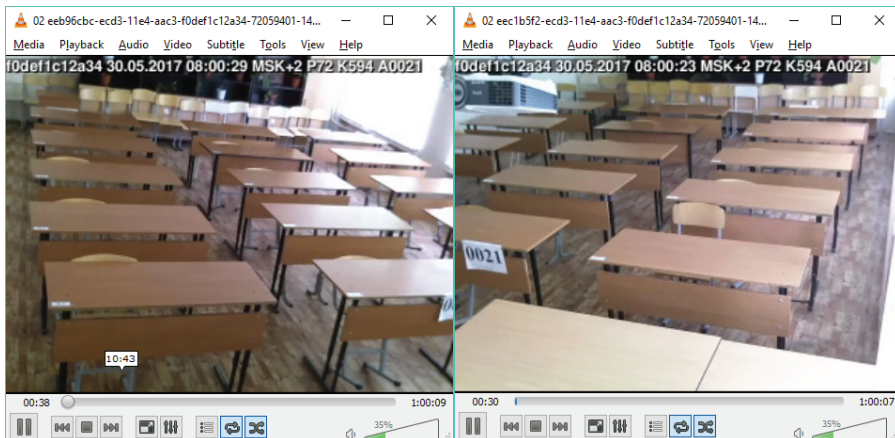


Рис. 1. Пример расположения камер в аудиториях при проведении экзамена

В ходе анализа нами было предложено представить каждого участника экзамена в виде сущности, отражающей следующую информацию о нем:

- 1) тип сущности – ученик, организатор или наблюдатель;
- 2) набор синтетических шаблонов, характерный для сущности определенного типа и позволяющий быстро идентифицировать участника экзамена на кадре.

Создание набора шаблонов базируется на идеях работы [3], в которой авторы строили синтетическую выборку на основе predetermined правил для распознавания дорожных знаков. При решении данной задачи мы осуществляем наблюдение за людьми, которые могут перемещаться по аудитории, совершать различные движения, не полностью попадать в кадр. Кроме того, набор шаблонов должен соответствовать сущностям определенного типа. Следовательно, искать вручную правила построения шаблонов затруднительно, и нами было предложено использовать для построения наборов шаблонов развернутые сверточные нейронные сети.

СПИСОК ЛИТЕРАТУРЫ

1. Wang Y. Spatiotemporal Pyramid Network for Video Action Recognition / Y. Wang, M. Long, J. Wang, Ph. S. Yu // 30th IEEE Conference on Computer Vision and Pattern Recognition. – Honolulu, Hawaii, 2017 – p. 2097-2107.
2. K. Simonyan Two-Stream Convolutional Networks for Action Recognition in Videos / K. Simonyan, A. Zisserman // Advances in Neural Information Processing Systems. – Vol. 27. – Montréal Canada, 2014.
3. Чигорин А. А., Конушин А. С. Эксперименты с обучением методов распознавания дорожных знаков на синтетических данных / А. А. Чигорин, А. С. Конушин // Наука и образование: научное издание МГТУ им. Н.Э. Баумана. – №8. – М.: Национальный Электронно-Информационный Консорциум, 2013 г. – стр. 315-324.