УДК 004.932

Программное средство идентификации скелета человека в видеопотоке

В.Н. Мымликов*, М.М. Фарафонов, П.В. Пересунько

Сибирский федеральный университет, пр. Свободный, 79,
Красноярск, 660041, Россия

*E-mail: vladmymlikov@mail.ru

Аннотация. Настоящая работа посвящена такому актуальному вопросу, как распознавание деятельности людей на видео. Исследование данного вопроса представляет большое научное и практическое значение для решения проблем обеспечения безопасности на предприятиях, а также корректности выполнения технологических операций на производстве. Оно необходимо при решении таких задач, как организация сетей видеонаблюдения в помещениях и на открытой местности. Поэтому необходимость создания IoT на основе распознавание деятельности людей на видео является важной задачей. Целью данной работы является исследование возможности применения методов машинного обучения, в частности CNN и LSTM сетей, для решения данной проблемы. В частности рассматривается возможность распознавания деятельности человека по перемещению ключевых точек тела. В данной работе приводится пример разработанной авторами системы, которая применяет упомянутые методики.

Ключевые слова: компьютерное зрение, распознавание деятельности человека, распознавание позы человека, LSTM

A software tool for identifying a human skeleton in a video stream

V.N. Mymlikov*, M.M. Farafonov, P.V. Peresunko

Siberian Federal University, 79 Svobodny pr., Krasnoyarsk, 660041, Russia

*E-mail: vladmymlikov@mail.ru

Abstract. This work is devoted to such a topical issue as the recognition of human activity in video. The study of this issue is of great scientific and practical importance for solving the problems of ensuring safety at enterprises, as well as the correctness of performing technological operations in production. It is necessary when solving problems such as organizing video surveillance networks in rooms and in open areas. Therefore, the need to create a software tool based on the recognition of human activity on video is an urgent task. The purpose of this work is to study the possibility of applying machine learning methods, in particular CNN and LSTM networks, to solve this problem. In particular, the possibility of recognizing human activity by moving key points of the body is considered. This paper provides an example of a system developed by the authors that applies the above techniques.

Keywords: computer vision, human activity recognition, human pose estimation, LSTM

1. Введение

Распознавание человеческой деятельности – это область исследований, связанных с идентификацией конкретного движения или действия человека на основе данных датчиков [1]. Классификация активности человека обычно производится на основании развернутой во времени информации. Существуют различные способы записи данных о движениях человека, это может быть видео, показания датчиков захвата движения [2] или же это могут быть показатели акселерометра и гироскопа смартфона [3]. Настоящая работа исследует возможность распознавания деятельности людей на видео со множества камер.

Один из методов классификации активности человека, заключается в анализе данных об изменении наклона его тела [4]. Однако у данного подхода есть недостатки [5].

Для классификации активности на видео можно применять сверточные нейронные сети – CNN, обученные распознавать вид деятельности по одному кадру.

Однако CNN потребуется большой набор данных. Другая проблема заключается в том, что обработка кадров по одному в отрыве от контекста приводит к неуверенности предсказания сети. Простое решение этой проблемы – усреднять прогнозируемые вероятности для n кадров. Кроме того, могут возникнуть проблемы с классификацией активностей, для которых важна последовательность действий. Разновидностями данного подхода являются методы позднего [6], [7] и раннего слияния [8], [9], [10].

Более совершенным решением является совместное использование CNN и LSTM сетей [11], [12]. Данный подход эффективнее простых сверточных сетей, благодаря возможности распознавать развернутую во времени информацию. Также существует метод, при котором сначала находятся точки тела человека, а затем эти точки передаются в сеть классификатора для определения активности [13], [14], [15].

2. Постановка задачи

В нашей работе мы выполняли классификацию 5-и классов действий: приседание, ходьба, сидение, стояние, неизвестный класс. Последний класс характеризует действия, которые не попадают под известные классы.

В качестве входных данных использовались последовательности кадров из видео файлов, записанных при помощи камеры. При этом камера была расположена на уровне торса и направлена горизонтально.

Первичными данными, выступили видеозаписи, на которых один человек, четко различимый на фоне окружения, выполняет определенное действие. В общей сложности было отснято 125 видеороликов длительностью 20 секунд каждый, таким образом, весь набор данных включает в себя примерно 41 минуту видео. На каждый класс действий приходится около 8 минут или 10000 кадров. Также был создан валидационный набор данных общей продолжительностью в 7 минут. На каждый класс приходится 50 секунд валидационных данных или около 1500 кадров.

3. Методы и материалы исследования

3.1. Способ получения скелета и трекинг людей

Для определения позы человека существует ряд готовых систем, задействующих машинное обучение и сверточные нейронные сети: OpenPose [16], AlphaPose [17], PoseFlow[18], DeepCut [19]. Главным преимуществом данных систем является их способность распознавать множество людей на одном кадре, что является нетривиальной задачей (рисунок 1). Мы считаем OpenPose более предпочтительным для нашей задачи, поскольку он быстрее при схожем уровне точности [16].

Рисунок 1. Примеры работы разных определителей поз.

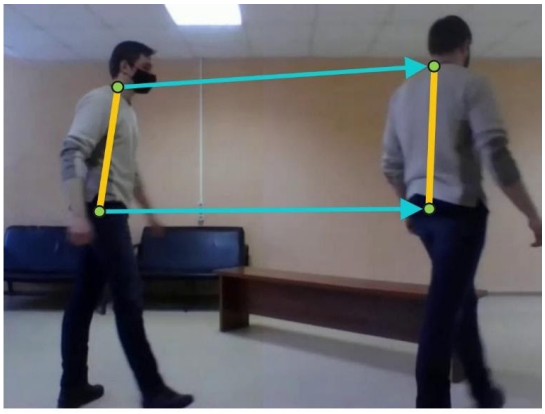

Рисунок 2. Принцип работы трекера.

Существуют различные алгоритмы и их модификаций, направленных на решение задачи трекинга [20], [21], [22]. К сожалению, имеющиеся решения не очень подходят для нашей системы. Простые алгоритмы быстры, но точность в них недостаточна для надежной классификации действий. Более продвинутые подходы точнее, но на получение результата уходит много времени, что затрудняет работу в реальном времени.

Реализованный в нашей работе алгоритм трекинга использует информацию о ключевых точках тела человека. В качестве метрики соответствия выступает разность между точками позвоночника скелета: чем меньше разность, тем больше вероятность того, что это один и тот же человек (рисунок 2).

Поскольку мы используем отдельную систему для оценки позы, то предложенное решение позволяет использовать полученные данные, которые в любом случае необходимы для классификации, и избежать необходимости повторной обработки кадра сторонним алгоритмом трекинга.

3.2. Способы нормализации скелетов, оба варианта

Предобработка данных состоит из двух этапов. Первый этап заключается в смещении скелета в начало координат, что нивелирует положение человека в кадре (рисунок 4). Второй этап заключается в нормализации координат точек скелета. Было рассмотрено два возможных варианта (рисунок 5).

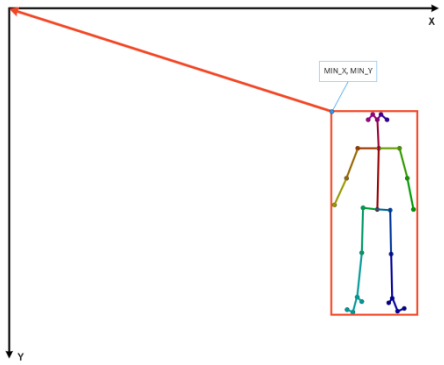

Рисунок 4. Первый этап нормализации данных.

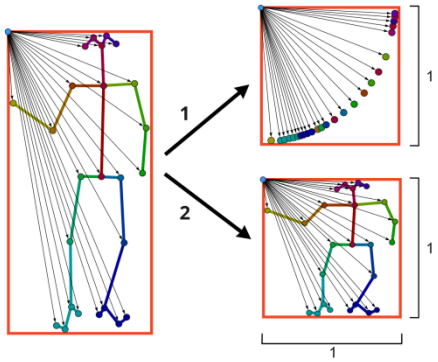

Рисунок 5. Варианты второго этапа нормализации.

Первый способ заключается в векторной нормализации каждой точки скелета, в результате чего они оказываются на дуге единичной окружности. Данный вариант работает, но в некоторых ситуациях теряется информация об относительном расположении точек скелета.

Второй способ состоит в том, чтобы нормировать точки по вектору с наибольшей длиной, что приводит к сжатию скелета до квадрата 1x1. При этом относительные расстояния между точками сохраняются.

3.3. Процесс обучения классификатора

В качестве классификатора действий применялась нейронная сеть с LSTM слоем. Для определения лучших параметров был использован метод перебора. В качестве настраиваемых параметров выступали: скорость обучения, способ ее изменения, размер мини-батча, оптимизатор, а также коэффициент дропаута. По итогам перебора были сделаны следующие выводы:

- использование сложной стратегии изменения скорости менее эффективно в сравнении с простым пошаговым изменением с шагом в 50 эпох и начальной скоростью обучения 0.005;
- размер мини-батча должен составлять ~700 примеров, значение менее 100 не позволяет сети обучиться в целом;
- Adam наиболее эффективная функция оптимизации в данном случае.

4. Полученные результаты

На графике (рисунок 6) приводится сравнение точности модели в зависимости от способа нормировки. При обучении сеть смогла показать точность на валидации ~91% к 325-ой эпохе (рисунок 7).

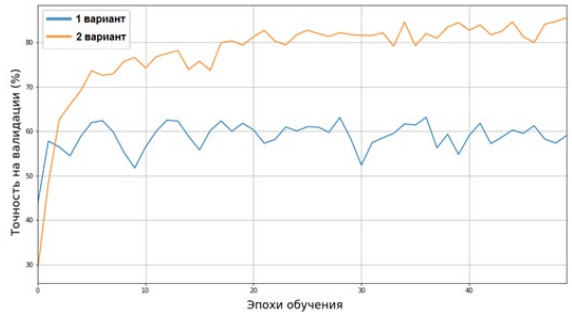

Рисунок 6. Точность при разных вариантах нормализации.

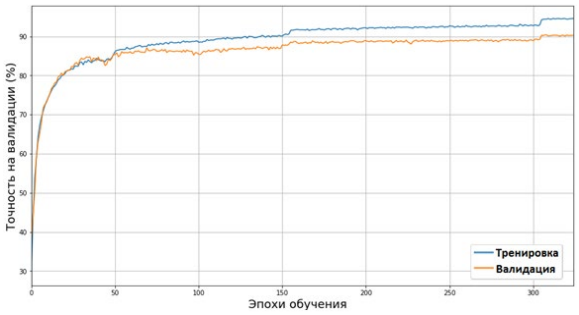

Рисунок 7. Итоговая точность системы.

Классификатор активности способен распознавать активность нескольких человек, принимая на вход множество видеопотоков от камер. Итоговый результат работы классификатора представлен ниже (рисунок 8).

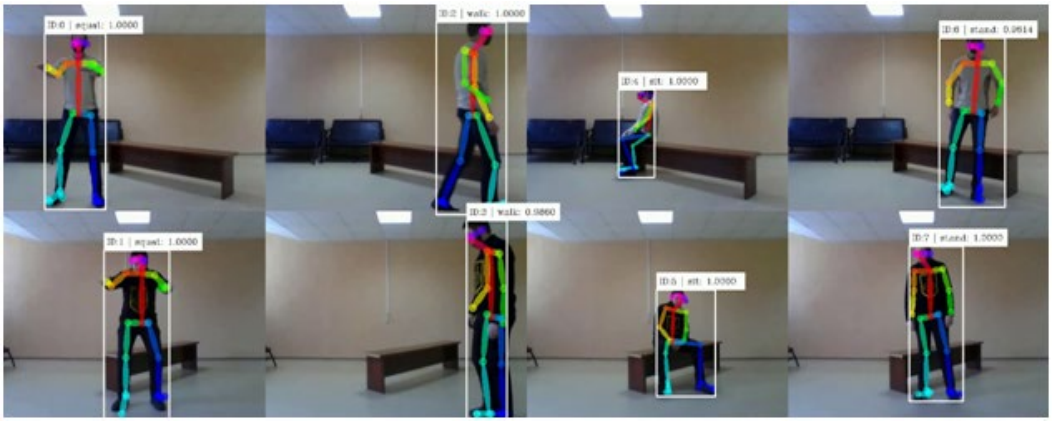

Рисунок 8. Результат работы системы.

5. Выводы

Разработанная система способна классифицировать 4 классов действий, в том числе приседания, стояние, сидение, ходьбу, а все остальное помечает как неизвестный класс. Разработанное решение показало итоговую точность около 91% правильно классифицированных кадров. Трекирование людей между кадрами осуществлялось при помощи оригинального метода – отслеживания позвоночника, который способен безошибочно отслеживать каждого человека на видео, в том числе в сложных ситуациях, например, когда люди проходят мимо друг друга. Однако есть еще большая область для дальнейшего исследования. В частности необходимо более подробно изучить возможность изменения числа кадров, по которым производится классификация, для распознавания более сложных и длительных действий. Также в целях улучшения точности распознавания скелета можно попробовать предварительно определять области, на которых присутствуют люди, и затем выполнять для них оценку позы и классификацию действий. Это имеет смысл, поскольку если человек расположен далеко от камеры, то при текущем подходе он может быть не распознан. Кроме того, перспективно выглядит идея продолжить настоящее исследования, но уже с распознаванием позы человека в трехмерном пространстве.

Список литературы

1. Browne, D. Deep Learning Human Activity Recognition / D. Browne // 27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science. – 2019. – P 76-87.
2. Mahmud, S. Human Activity Recognition from Wearable Sensor Data Using Self-Attention / S. Mahmud, T. Tonmoy, K. Bhaumik, M. Rahman, A. Amin, M. Shoyaib, M. Khan, A. Ali // 24th European Conference on Artificial Intelligence. – 2020. – P 1332-1339.

3. Baloch, Z. Deep Architectures for Human Activity Recognition using Sensors / Z. Baloch, F. Shaikh, M. Una // 3C Tecnologia. Special Issue. – 2019. – № 29-2. – P 15-36.
4. Davide, A. A public domain dataset for human activity recognition using smartphones / A. Davide, G. Alessandro, O. Luca, P. Perez, X. Ortiz, J. Luis // 21st European Symposium on Artificial Neural Networks, Computational Intelligence And Machine Learning. – 2013. – P 24-26.
5. Tong, C. Are Accelerometers for Activity Recognition a Dead-end? / C. Tong, S. Tailor S, N. Lane // HotMobile '20: The 21st International Workshop on Mobile Computing Systems and Applications. – 2020. – P 1-6.
6. Dhiman, C. View-invariant deep architecture for human action recognition using two-stream motion and shape temporal dynamics / C. Dhiman, D. Vishwakarma // IEEE Transactions on Image Processing. – 2020. – № 29 – P 3835-3844.
7. Jalal, M.A. Dual stream spatio-temporal motion fusion with self-attention for action recognition / M.A. Jalal, W. Aftab, R.K Moore, L. Mihaylova // 2019 22th International Conference on Information Fusion (FUSION). 22nd International Conference on Information Fusion. – 2019. – P 1-8.
8. Kumar, P. Human activity recognition with deep learning: overview, challenges and possibilities / P. Kumar, S. Chauhan // CCF Trans. Pervasive Comp. Interact. – 2021. – P 1-21.
9. Serpush, F. Complex Human Action Recognition in Live Videos Using Hybrid FR-DL Method / F. Serpush, M. Rezaei // arXiv preprint arXiv:2007.0281.1 – 2020. – P. 1-14.
10. Das, S. A new hybrid architecture for human activity recognition from rgb-d videos / S. Das, M. Thonnat, K. Sakhalkar, M. Koperski, F. Bremond, G. Francesca // International Conference on Multimedia Modeling – 2019 – P 493-505.
11. Ullah, A. Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features / A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, S. Baik // IEEE access. – 2017. – № 6. – P 1155-1166.
12. Aziz, K. An Efficient Human Activity Recognition Technique Based on Deep Learning / K. Aziz, F. Ababsa, N. Benoudjit // Pattern Recognition and Image Analysis. – 2019. – 29(4). – P 702-715.
13. Gupta, A. Human Activity Recognition Using Pose Estimation and Machine Learning Algorithm / A. Gupta, K. Gupta, K. Gupta, K. Gupta // ISIC'21: International Semantic Intelligence Conference. – 2021. – P 25-27.

14. Luvizon, D. 2d/3d pose estimation and action recognition using multitask deep learning / D. Luvizon, D. Picard, H. Tabia // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition – 2018. – P. 5137-5146.
15. Yao, A. Coupled action recognition and pose estimation from multiple views / A. Yao, J. Gall, L. Van Gool // International journal of computer vision. – 2012. – P. 16-37.
16. Cao, Z. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields / Z. Cao, G. Hidalgo, T. Simon, SE. Wei, Y. Sheikh, 2019. – P. 1-14. arXiv:1812.08008v2.
17. Fang, HS. RMPE: Regional Multi-person Pose Estimation / H.S. Fang, S. Xie, Y.W. Tai, C. Lu, 2018. – P. 1-10. arXiv:1612.00137v5.
18. Xiu, Y. Pose Flow: Efficient Online Pose Tracking / Y. Xiu, J. Li, H. Wang, Y. Fang, C. Lu, 2018. – P. 1-12. arXiv:1802.00977v2.
19. Pishchulin, L. DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation / L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. Gehler, B. Schiele, 2016. – P. 1-15. arXiv:1511.06645v2.
20. Bewley, A. Simple Online and Realtime Tracking / A. Bewley, Z. Ge, L. Ott, F. Ramos, B. Upcroft, 2017. – P. 1-5. arXiv:1602.00763v2.
21. Kim, C. Multiple Hypothesis Tracking Revisited / C. Kim, F. Li, A. Ciptadi, J. M. Rehg // 2015 IEEE International Conference on Computer Vision (ICCV). – 2015. – P. 4696-4704. 10.1109/ICCV.2015.533.
22. Xiu, Y. Pose Flow: Efficient Online Pose Tracking / Y. Xiu, J. Li, H. Wang, Y. Fang, C. Lu, 2018. – P. 1-12. arXiv:1802.00977v2.