УДК УДК 004.93'1

DOI: 10.15827/0236-235X.129.162-168

Дата подачи статьи: 23.07.19

2020. Т. 33. № 1. С. 162–168

Интеллектуальный анализ видеоданных для распознавания ситуаций угона автомобиля на парковке

А.Ю. Кручинин¹, к.т.н., доцент кафедры ВТuЗИ ОГУ, kruchinin-al@mail.ru

Р.Р. Галимов¹, к.т.н., доцент, rin-galimov@yandex.ru

¹ Оренбургский государственный университет, г. Оренбург, 460018, Россия

Для предотвращения и расследования различных инцидентов активно используются системы видеонаблюдения. Основным недостатком традиционных решений в этой области является то, что оценивает текущую ситуацию и принимает решение непосредственно оператор, поэтому при большом количестве контролируемых камер и в силу человеческого фактора существует высокая вероятность увеличения времени определения опасной ситуации. Это приводит к значительному ущербу, особенно в тех случаях, когда необходимо оперативно реагировать на инцидент, например, при угона автомобильного транспорта. Данные обстоятельства обуславливают необходимость внедрения систем интеллектуального анализа видеоданных.

В работе предложен метод распознавания ситуаций угона автомобильного транспорта с парковки, основанный на стохастических грамматиках и глубоких нейронных сетях. Распознавание инцидента происходит на двух основных уровнях: на нижнем уровне распознаются события, а на верхнем – ситуации как наиболее вероятные цепочки событий, соответствующие грамматике сигнатур угона автомобиля. Детектирование событий возможного угона выполняется на основе анализа выявленных объектов, их взаимного расположения, динамических характеристик траектории движения и особенностей поз людей. Распознавание объектов и поз людей осуществляется на основе глубоких нейронных сетей, характеризующихся на современном этапе развития высокой степенью достоверности.

В статье описана разработанная имитационная модель системы распознавания ситуаций угона автомобильного транспорта, которая базируется на модуле распознавания объектов с помощью глубокой нейронной сети. Повышение оценки достоверности распознавания события осуществляется за счет учета истории определенных объектов на предыдущих кадрах и при необходимости данных о позе человека.

Для описания возможных сценариев угона автомобильного транспорта разработана стохастическая грамматика, на основе которой создана тестовая утилита. Результаты тестирования разработанного метода на наборе данных Mini-deone video dataset показали ее работоспособность.

Ключевые слова: распознавание угона, глубокая нейронная сеть, стохастические грамматика, событийно-ситуационный подход, распознавание позы человека.

Активное развитие аппаратных средств систем видеонаблюдения привело к увеличению количества используемых видеокамер в различных областях. Увеличение объема видеоданных в системах видеонаблюдения для предотвращения противоправных действий требует автоматизации процесса распознавания сложных ситуаций. Этим обусловлена необходимость разработки методов и средств автоматизированного распознавания сложных ситуаций, позволяющих своевременно определять нарушения и сообщать о них.

Одним из типов правонарушений, на которые нужно своевременно реагировать, являются угоны автомобильных транспортных средств. Цель авторов данной работы – автоматизация распознавания ситуаций угона автомо-

билей на основе событийно-ситуационного подхода.

Событийно-ситуационный подход к распознаванию образов

В работе [1] предлагается подход к событийно-ситуационному распознаванию, в котором выделяются этапы распознавания объектов, событий и ситуаций. Объект – это графический образ (человек, автомобиль, идентификационный номер). Событие – мгновенное совершение действия объектом или без объекта (пересечение линии, возникновение огня). Ситуация – последовательность событий (например, человек пришел, оставил сумку и ушел).

Для распознавания образов очень эффективным с точки зрения достоверности является использование различных классов глубоких нейронных сетей [2]. Применительно к системе видеомониторинга эти задачи распознавания можно разделить на несколько групп:

- задачи распознавания объектов: распознавание изображения в целом, сегментация изображения, детектирование отдельных объектов, детектирование объекта и его структуры (например, скелета объекта);
- задачи распознавания событий: наблюдение за изменением изображения в целом, распознавание изменения позиции/состояния детектированного объекта;
- задачи распознавания ситуаций: анализ последовательности событий с одним объектом, взаимодействия отдельных объектов между собой, последовательности изменения изображения в целом.

Распознавание изображений в целом и детектирование отдельных объектов решаются сетями AlexNet, ZF Net, VGG Net, ResNet-50, YOLO, GoogLeNet и т.п. Для сегментации изображений существуют такие сети, как SegNet, FSN, ParseNet, U-Net, FPN, PSPNet, R-CNN и другие. Распознавание структуры объекта – сложная задача. Например, фреймворк OpenPose предназначен для распознавания позы человека (скелета), в основе которого лежит многоступенчатая модель Multi-Person Pose Estimation [3].

Для решения задачи, связанной с распознаванием образов при угоне автомобиля, достаточно использовать модели, которые с необходимой точностью распознают автомобили, людей и их позы. В данной работе были использованы нейронная сеть YOLO [4], фреймворк tf-pose-estimation [5], а также стохастические грамматики [6].

В работе [1] предложена имитационная модель распознавания ситуаций в реальном времени. Руководствуясь ею, была построена имитационная модель распознавания ситуаций с угоном автомобилей на парковке (рис. 1).

Источник видеопотока V выдает периодически изображения, которые необходимо распознать. Текущее изображение I_x поступает на начальную обработку – распознавание образов средствами YOLO. Блок распознавания образов выдает множество распознанных графических образов $\Omega_x = \{\omega_{x1}, \omega_{x2}, \dots, \omega_{xL}\}$, которые поступают на блок распознавания событий. Распознавание позы человека по запросу z выделено в отдельный блок для снижения нагрузки на вычислительные ресурсы. Распознавание в этом блоке вызывается в случае, когда есть признаки опасного события, например, человек находится около автомобиля. В этом случае возвращается информация о его позе Pos_x . Далее аналогично работе [1] блок распознавания событий передает Ω_x в базу предыдущих образов, из которой последовательность ранее появлявшихся образов $\Omega = \{\Omega_1, \Omega_2, \dots$,

Рис. 1. Имитационная модель распознавания ситуаций с угоном автомобиля на парковке

Fig. 1. Simulation model of recognizing situations with car theft in the parking lot

Ω_x возвращается обратно в блок распознавания событий. На выходе блока распознавания событий выдается идентифицированное событие E_x (E_x может быть пустым, что означает отсутствие событий), которое поступает на блок распознавания ситуаций. Блок распознавания ситуаций передает E_x в базу предыдущих событий, из которой возвращается последовательность событий E . На выходе блока распознавания ситуаций – распознанная ситуация S_x , которая является выходом модели и передается на блок расчета управляющего воздействия.

В работе [7] предложен подход к распознаванию ситуаций на основе стохастических грамматик. В данной работе рассматриваются две ситуации угона: 1) человек подходит к машине, долго стоит или присаживается, пытается открыть дверь, садится в машину и уезжает; 2) человек подбегает к другому человеку, открывающему автомобиль, отталкивает его, садится в машину и уезжает.

Вероятность возникновения опасной ситуации (угона автомобиля) можно выразить следующей простой грамматикой в форме Бекуса–Наура: $G = (N, T, P, S)$, где N – множество не терминальных (служебных) символов; T – множество терминальных символов; P – правила вида $\alpha \rightarrow \beta$ (α, β – цепочки); S – начальный символ. Можно выделить следующие события, распознаваемые при угона автомобилей в описанных выше сценариях:

- человек появляется в области слежения, a_0 ;
- человек идет, a_1 ;
- человек бежит, a_2 ;
- человек находится около автомобиля, a_3 ;
- человек стоит около автомобиля с руками на уровне талии, направленными вперед, в течение некоторого времени (например, от 10 секунд), a_4 ;
- человек присел около автомобиля с руками на уровне головы, направленными вперед, a_5 ;
- один человек бежит к другому, стоящему около машины или севшему в машину, a_6 ;
- контакт людей, a_7 ;
- человек движется от автомобиля в неестественной позе, a_8 ;
- человек садится в машину, a_9 ;
- машина уезжает, a_{10} ;
- неизвестное событие, a_{11} .

Грамматика будет иметь следующий вид:

$G = (\{S, A, B, C, D, E, F, M, H, I, J, K, L\}, \{a_0, a_1,$

$a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}\}, P, S)$, где P : $S \rightarrow a_0AB$, $A \rightarrow a_1/a_2$, $B \rightarrow a_3C/AB/a_{11}$, $C \rightarrow DE/a_9F/AB/a_6M$, $D \rightarrow a_4/a_5$, $E \rightarrow a_9F/AB$, $F \rightarrow a_6J/a_{11}C/a_{10}$, $M \rightarrow a_7K/a_{11}F$, $H \rightarrow a_8I/a_{11}C$, $I \rightarrow a_9a_{10}/a_{11}H$, $J \rightarrow a_7H/a_{11}C$, $K \rightarrow a_8L/a_{11}F$, $L \rightarrow a_9a_{10}/a_{11}K$.

Конечный автомат для грамматики представлен в форме графа на рисунке 2. Начальное состояние – 0, конечные состояния – 5, 12.

Рис. 2. Конечный автомат для грамматики событий, описывающих угон автомобиля

Fig. 2. State machine for the grammar of events that determine the car theft

Приведем цепочки событий, соответствующие вероятному угону автомобиля:

$\mu_1: a_0a_1a_3a_6a_7a_8a_9a_{10}$

$\mu_2: a_0a_1a_3a_9a_6a_7a_8a_9a_{10}$

$\mu_3: a_0a_1a_3a_4a_9a_{10}$

$\mu_4: a_0a_1a_3a_5a_9a_{10}$

Остальные случаи рассматриваем как ситуации, не соответствующие угону.

Несмотря на достоинства методов распознавания образов на основе глубоких нейронных сетей, результаты формируются с определенной степенью достоверности. Особенно это касается событий, в которых важно определение положения частей тела человека. Так, в [8–10] показано, что поза человека определяется с некоторой степенью достоверности. Например, события a_4 или a_5 достаточно сложно распознать с высокой степенью достоверности, поэтому возможны альтернативные варианты, в частности, для цепочки μ_3 : $\mu_5: a_0a_1a_3a_9a_{10}$.

Таким образом, возникает задача определения наиболее вероятной цепочки из всех возможных. Оценки вероятности конкретной ситуации рассчитываются на основе формулы Байеса.

Для задачи распознавания угона автомобильного транспортного средства определены распределения вероятностей по правилам грамматики G :

$S \xrightarrow{1.0} a_0 AB; A \xrightarrow{0.5} a_1; A \xrightarrow{0.5} a_2;$
 $B \xrightarrow{0.4} a_3 C; B \xrightarrow{0.4} AB; B \xrightarrow{0.2} a_{11};$
 $C \xrightarrow{0.2} DE; C \xrightarrow{0.3} a_9 F; C \xrightarrow{0.3} AB;$
 $C \xrightarrow{0.2} a_6 M;$
 $D \xrightarrow{0.5} a_4; D \xrightarrow{0.5} a_5;$
 $E \xrightarrow{0.6} a_9 F; E \xrightarrow{0.4} AB;$
 $F \xrightarrow{0.2} a_6 J; F \xrightarrow{0.1} a_{11} C; F \xrightarrow{0.7} a_{10};$
 $J \xrightarrow{0.4} a_7 H; J \xrightarrow{0.6} a_{11} C;$
 $H \xrightarrow{0.3} a_8 I; H \xrightarrow{0.7} a_{11} C;$
 $I \xrightarrow{0.9} a_9 a_{10}; I \xrightarrow{0.1} a_{11} H;$
 $M \xrightarrow{0.4} a_7 K; M \xrightarrow{0.6} a_{11} F;$
 $K \xrightarrow{0.3} a_8 L; K \xrightarrow{0.7} a_{11} F;$
 $L \xrightarrow{0.4} a_9 a_{10}; L \xrightarrow{0.6} a_{11} K.$

Для распознавания ситуаций угона разработано тестовое программное приложение DetectCarTheft, структурная схема которого представлена на рисунке 3.

Модуль управления инициализирует процесс распознавания ситуации угона автомобиля на парковке, формируя сигналы St_1 и St_2 соответственно модулям захвата видео и распознавания ситуаций. Обмен информацией между элементами программного средства производится через модуль данных, который

хранит следующие типы данных: текущий кадр I , набор распознанных событий E , данные о распознанных в текущем кадре событиях Ed .

Важным элементом данной системы является модуль обработки событий, который осуществляет детектирование события на основе результата обнаружения объектов с помощью глубокой нейронной сети, выявления особенности траектории и определения взаимного расположения. Для повышения достоверности распознавания текущего события E_t используются данные предыдущих результатов E_{t-N} с длительностью истории N .

При необходимости модуль обработки событий дополнительно запрашивает информацию о позе человека, например, если распознанный контур человека находится вблизи автомобиля. На основе распознанных событий модуль распознавания ситуаций осуществляет разбор цепочек и определяет наиболее вероятную ситуацию. Промежуточные результаты работы программы представлены на рисунке 4. Использовался тестовый набор данных Mini-drone video dataset [11].

Для рассматриваемого примера была выявлена цепочка событий μ_1 : $a_0, a_1(0.95), a_3, a_5(0.9), a_9, a_{10}$.

Альтернативные варианты цепочки ситуаций:

μ_2 : $a_0, a_2, a_3, a_5, a_9, a_{10}$

μ_3 : $a_0, a_1, a_3, a_4, a_9, a_{10}$

μ_4 : $a_0, a_2, a_3, a_4, a_9, a_{10}$.

Рис. 3. Структурная схема программы DetectCarTheft

Fig. 3. The structural scheme of the program DetectCarTheft

165

Данные цепочки событий соответствуют разработанной грамматике G угона автомобиля. С учетом подхода, представленного

в работе [7], осуществлена оценка вероятности такой ситуации, которая составляет 85,5 %.

Заключение

Разработанный метод распознавания ситуации угона позволяет повысить оперативность обнаружения противоправного действия за счет автоматизации данного процесса. Высокая достоверность определяется двумя основными факторами: достаточно высоким качеством распознавания объектов и их местоположения глубокими нейронными сетями и

задания сигнатуры угрозы на основе стохастической грамматики.

Данный подход имеет перспективы применения и в других областях, но требует качественной разработки распознаваемых грамматик. Кроме того, необходимо отметить, что предложенный подход легко может быть модифицирован для применения в распределенных системах видеонаблюдения с несколькими камерами.

Литература

1. Колмыков Д.В., Кручинин А.Ю. Имитационная модель распознавания ситуаций на основе структурных методов в системах видеонаблюдения // Информационные системы и технологии. 2017. № 2. С. 25–31.
2. Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, Xindong Wu. Object detection with deep learning: a review. URL: <https://arxiv.org/pdf/1807.05511.pdf> (дата обращения: 20.07.2019).
3. Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: Realtime Multi-person 3D Pose Estimation Using Part Affinity Fields. In arXiv preprint arXiv:1812.08008, 2018.
4. Yolo-v3 and Yolo-v2 for Windows and Linux. URL: <https://github.com/AlexeyAB/darknet> (дата обращения: 20.07.2019).
5. Deep Pose Estimation Implemented using Tensorflow with Custom Architectures for Fast Inference. URL: <https://github.com/ildoonet/tf-pose-estimation> (дата обращения: 20.07.2019).
6. Фу К. Структурные методы в распознавании образов; [пер. с англ.]. М.: Мир, 1977. 319 с.
7. Галимов Р.Р., Кручинин А.Ю. Структурное распознавание в системах видеонаблюдения на основе стохастических грамматик // Информационные системы и технологии. 2018. № 5. С. 15–21.
8. Yang X., Tian Y. Eigenjoints-based action recognition using naive-bayes-nearest-neighbor. Proc. Work. on Human Activity Understanding from 3D Data, Providence, Rhode Island, 2012, pp. 14–19.
9. Xia L., Chen C., Aggarwal J.K. View invariant human action recognition using histograms of 3D joints. Proc. CVPRW, Providence, Rhode Island, 2012, pp. 20–27. DOI: 10.1109/CVPRW.2012.6239233.
10. Devanne M., Wannous H., Berretti S., Pala P., Daoudi M., Del Bimbo A. 3D human action recognition by shape analysis of motion trajectories on riemannian manifold. Transactions on Systems Man and Cybernetics, 2015, vol. 7, no. 45, pp. 1340–1352. DOI: 10.1109/TCYB.2014.2350774.
11. Mini-drone Video Dataset. URL: <https://mmspg.epfl.ch/downloads/mini-drone/> (дата обращения: 20.07.2019).

Software & Systems
DOI: 10.15827/0236-235X.129.162-168

Received 23.07.19
2020, vol. 33, no. 1, pp. 162–168

Intelligent analysis video data to recognize car theft situations in a parking lot

A.Yu. Kruchinin¹, Ph.D. (Engineering), Associate Professor of the Department of Computer Science and Information Protection, kruchinin-al@mail.ru

R.R. Galimov¹, Ph.D. (Engineering), Associate Professor, rin-galimov@yandex.ru

¹ Orenburg State University, Orenburg, 460018, Russian Federation

Abstract. Modern video surveillance systems are very popular to prevent and investigate various illegal incidents. The main disadvantage of traditional decisions in this area is that the operator directly assesses the current situation and makes a decision. With a large number of monitored cameras and due to the human factor, there is a high probability of increasing the time delay in determining the dangerous situation. This leads to significant damage, especially in cases where it is necessary to respond promptly to an incident. Such cases include, for example, the situation of carjacking. These circumstances necessitate the introduction of systems for the intellectual analysis of video data.

This paper submits a method for recognizing situations of car theft from parking based on stochastic grammars and deep neural networks. Recognition of an incident is at two main levels: at the lower level, event recognition occurs, at the upper level, the situation as the most probable chain of events corresponding to the grammar of car theft signatures. The analysis of the identified objects, their relative position, the dynamic characteristics of the trajectory of movement and the characteristics of the people's postures, fills up the detection of possible hijacking events. The deep neural networks recognize the objects and people poses. These networks have a high degree of reliability.

The article developed a simulation model of the system of recognition of the situations of theft of road transport, which based on the object detection module using a deep neural network. The history of certain objects in previous frames and, if necessary, data on the posture of a person improves the assessment of the reliability of event recognition.

To describe the possible scenarios of car theft, the authors developed a stochastic grammar and created the test utility based on it. The test results of the developed method on the Mini-deone video dataset showed its efficiency.

Keywords: hijacking recognition, deep neural network, stochastic grammars, event-situational approach, and human posture recognition.

References

1. Kolmykov D.V., Kruchinin A.Yu. Simulation model of the situations recognition on the basis of structural methods surveillance. *Information Systems and Technologies*. 2017, no. 2, pp. 25–31 (in Russ.).
2. Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, Xindong Wu. *Object Detection with Deep Learning: A Review*. Available at: <https://arxiv.org/pdf/1807.05511.pdf> (accessed July 20, 2019).
3. Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. *OpenPose: Realtime Multi-person 3D Pose Estimation Using Part Affinity Fields*. In arXiv preprint arXiv:1812.08008, 2018.
4. *Yolo-v3 and Yolo-v2 for Windows and Linux*. Available at: <https://github.com/AlexeyAB/darknet> (accessed July 20, 2019).
5. *Deep Pose Estimation Implemented Using Tensorflow with Custom Architectures for Fast Inference*. Available at: <https://github.com/ildoonet/tf-pose-estimation> (accessed July 20, 2019).
6. Fu K.S. *Syntactic Methods in Pattern Recognition*. 1974, 306 p. (Russ. ed.: Moscow, 1977, 319 p.).
7. Galimov R.R., Kruchinin A.Yu. Structural recognition in video surveillance systems based on stochastic grammars. *Information Systems and Technologies*, 2018, no. 5, pp. 15–21.
8. Yang X., Tian Y. Eigenjoints-based action recognition using naive-bayes-nearest-neighbor. *Proc. Work. on Human Activity Understanding from 3D Data*. Providence, Rhode Island, 2012, pp. 14–19.
9. Xia L., Chen C., Aggarwal J.K. View invariant human action recognition using histograms of 3D joints. *Proc. CVPRW*. Providence, Rhode Island, 2012, pp. 20–27. DOI: 10.1109/CVPRW.2012.6239233.
10. Devanne M., Wannous H., Berretti S., Pala P., Daoudi M., Del Bimbo A. 3D human action recognition by shape analysis of motion trajectories on riemannian manifold. *Transactions on Systems Man and Cybernetics*. 2015, vol. 7, no. 45, pp. 1340–1352. DOI: 10.1109/TCYB.2014.2350774.
11. *Mini-drone Video Dataset*. Available at: <https://mmspg.epfl.ch/downloads/mini-drone/> (accessed July 20, 2019).

Для цитирования

Кручинин А.Ю., Галимов Р.Р. Интеллектуальный анализ видеоданных для распознавания ситуаций угона автомобиля на парковке // Программные продукты и системы. 2020. Т. 33. № 1. С. 162–168. DOI: 10.15827/0236-235X.129.162-168.

For citation

Kruchinin A.Yu., Galimov R.R. Intelligent analysis video data to recognize car theft situations in a parking lot. *Software & Systems*. 2020, vol. 33, no. 1, pp. 162–168 (in Russ.). DOI: 10.15827/0236-235X.129.162-168.