

Применение метода детектирования границ к задаче распознавания обстановки

А.А.Карандеев^{1,2}, В.П.Осипов¹, В.И.Балута^{1,2}

¹ ИПМ им. М.В.Келдыша РАН, Миусская пл, д.4, Москва, 125047, Россия

² РЭУ им. Г.В. Плеханова, Стремянный пер., 36, Москва, 117997, Россия

Аннотация

В настоящей работе представлены результаты одного из вариантов решения проблемы повышения скорости принятия решений интеллектуальным агентом при моделировании поведения сложных систем на виртуальном электронном полигоне. Такой полигон в настоящее время рассматривается в качестве инструментальной площадки для отработки технологий обучения интеллектуальных агентов в условиях обстановки различной сложности с целью последующего переноса отработанных методов на реальные объекты для решения практических задач. В качестве примера рассмотрено управление робототехническим средством, функционирующим в закрытом помещении. В статье изложена технология снижения объема и размерности обрабатываемых данных с целью повышения оперативности реагирования на изменение обстановки и выработку решений на перемещение робототехнического устройства. В основе технологии лежит предобработка видеоизображений для формирования обучающей выборки, а также процедура и результаты глубокого обучения сверточной нейросети. В работе использована библиотека алгоритмов компьютерного зрения OpenCV с открытым исходным кодом, реализованная на C/C++. Показано, что акцентирование внимания на выделении границ объектов позволяет существенно снизить объем данных для анализа ситуации и повысить скорость принятия решений роботом на перемещение.

Ключевые слова

Интеллектуальные агенты, повышение оперативности выработки решений, сверточная нейронная сеть, обработка видеоизображений

Application of the boundary detection method to the problem of recognizing the situation

A.A.Karandeev^{1,2}, V.P.Osipov¹, V.I.Baluta^{1,2}

¹ Keldysh Institute of Applied Mathematics, Miuskaya sq., 4, Moscow, 125047, Russia

² Plekhanov Russian University of Economics, Stremyanyi per., 36, Moscow, 117997, Russia

Abstract

This paper presents the results of one of the solutions to the problem of increasing the speed of decision-making by an intelligent agent when modeling the behavior of complex systems on a virtual electronic polygon. Such a training ground is currently considered as an instrumental platform for testing technologies for training intelligent agents in conditions of varying complexity in order to subsequently transfer the developed methods to real objects for solving practical problems. As an example, the control of a robotic device operating in an enclosed space is considered. The article describes the technology of reducing the volume and dimension of the processed data in order to increase the responsiveness to changes in the situation and the development of solutions for moving a robotic device. The technology is based on the preprocessing of video images for the formation of a training sample, as well as the procedure

ГрафиКон 2021: 31-я Международная конференция по компьютерной графике и машинному зрению, 27-30 сентября 2021 г., Нижний Новгород, Россия

EMAIL: karalex755@gmail.com (А.А.Карандеев); osipov@keldysh.ru (В.П.Осипов); vbaluta@keldysh.ru (В.И.Балута)

ORCID: 0000-0001-8555-996X (А.А.Карандеев); 0000-0002-8237-1053 (В.П.Осипов); 0000-0002-4399-0943 (В.И.Балута)

and results of deep learning of a convolutional neural network. The paper uses an open source library of OpenCV computer vision algorithms implemented in C / C++. It is shown that focusing on the selection of object boundaries can significantly reduce the amount of data for analyzing the situation and increase the speed of decision-making by the robot to move.

Keywords

Intelligent agents, improving the efficiency of decision-making, convolutional neural network, video image processing

1. Введение

Робототехника – прикладная наука, занимающаяся разработкой автоматизированных технических систем, является одним из передовых и наиболее стремительно развивающихся направлений прикладной математики и искусственного интеллекта. Перед ней стоит много задач, многие из которых еще только предстоит решить в дальнейшем, однако уже сейчас использование робототехники и роботов, в частности, дает свои положительные плоды.

Прикладная робототехника находит свое применение и в различных сферах научной деятельности, в частности, при решении задач компьютерного и математического моделирования различных процессов. В решении подобного рода задач робототехнические средства обычно выступают в роли агентов, а взаимодействуя вместе с другими роботами в рамках одного полигона образуют мультиагентные системы.

Предварительное описание подходов к построению электронного полигона в агентном представлении было сделано нами ранее в работах [1, 2] применительно к задаче конфликтного взаимодействия двух противодействующих сторон. Технологии агентного представления взаимодействующих элементов сложных систем в последние годы получили интенсивное развитие прежде всего благодаря возможности независимого описания сложного поведения каждого такого элемента, что, с одной стороны, существенно упрощает модельную постановку имеющихся задач при одновременном сохранении вариативности сложного поведения состоящей из таких элементов системы, с другой стороны, позволяет исследовать сложное поведение больших систем, которое невозможно реализовать другими методами.

При агентном представлении сложных систем одной из ключевых задач моделирования является задача математического представления интеллектуальных агентов [3-5].

Алгоритм поведения интеллектуального агента в процессе взаимодействия с окружающей средой можно представить в виде циклической последовательности шагов, представленных на рисунке 1. В соответствии с имеющейся у него картиной мира и целевой установкой агент оценивает обстановку, прогнозирует ее развитие, выбирает из возможных вариантов действий наиболее рациональное в плане приближения к цели и осуществляет его, после чего собирает информацию об изменении ситуации, проверяет эффективность совершенного действия, при необходимости корректирует свою модель представлений об окружающем мире и с учетом этого вновь оценивает изменившуюся обстановку.



Рисунок 1: Циклический алгоритм поведения интеллектуального агента

Логика поведения агента подобна логике поведения человека, но задается с помощью функции агента, которая отображает любую конкретную последовательность актов восприятия на некоторое действие среды [6].

Важнейшей особенностью интеллектуального агента является способность функционировать в условиях неточной, нечеткой и даже противоречивой информации, поскольку в силу ограниченных возможностей рецепторов и эффекторов агента он не может ни определять исчерпывающим образом параметры среды, ни точно предсказывать результаты намеченного действия, особенно в динамично изменяющейся обстановке.

Несмотря на схожесть логики принятия решений интеллектуальным агентом с человеческой, существует ряд принципиальных отличий между восприятиями обстановки человеком и компьютерными системами. Известно, что человек мыслит образами, сопоставляя то, что он видит и ощущает, с уже имеющимся у него опытом. Что, в свою очередь, при наличии даже неполной информации, позволяет достраивать цельную картину мира на основе прошлого опыта, выделять в этой картине существенные для выработки решения аспекты, принимать решение практически мгновенно [7]. В то время как компьютерные системы мыслят «конкретными данными». Это обстоятельство порождает определенное противоречие. С одной стороны, для повышения качества выработки рациональных решений необходимо максимально расширять объем имеющейся информации. С другой стороны, как попытки увеличить этот объем путем сбора дополнительной информации, так и обработка расширенного объема данных существенно замедляют процессы получения тех самых решений. Что может приводить к потере актуальности выработанного решения в связи с изменением обстановки [8, 9]. Выигрыш в качестве решений приводит к проигрышу по времени их получения. В условиях электронного полигона, где происходит многократное (сотни тысяч и более) моделирование взаимодействия множества агентов, такое положение неприемлемо. В связи с чем возникает проблема совершенствования алгоритмов формирования интеллектуальным агентом его картины мира с возможностью выделения только существенных деталей с последующей оценкой обстановки на основе такой неполной информации без потери адекватности.

2. Описание технологии

Задача оценки обстановки в значительной мере аналогична задаче распознавания образов [10-12]. Опираясь на эту аналогию, обработка основных подходов к решению сформулированной проблемы была проведена на примере робототехнических средств с использованием реальных данных, полученных путем видео и фотосъемки роботизированной платформой «PATROLBOT», базовая версия которой представлена на рисунке 2. Данная платформа является мобильной, оснащена рядом сенсорных устройств, может работать автономно в течение длительного периода, решая задачи патрулирования и доставки грузов, причем у оператора есть возможность просматривать в реальном времени получаемые с сенсоров данные, включая видеопотоки.

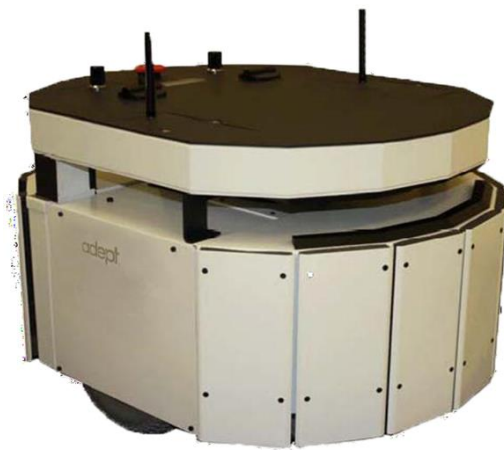


Рисунок 2: Платформа робота-помощника «PATROLBOT»

Хотя для ориентации и позиционирования в пространстве платформой применяется ряд сонаров, лазерных дальномеров и позиционных датчиков, в рамках решаемой в настоящей работе задачи для классификации объектов и окружающей обстановки использовались только видеоизображения, получаемые стереопарой.

Основная задача заключалась в разработке программного комплекса для формирования датасета и последующего обучения нейронной сети совместно с методами ускоренной оценки обстановки с целью оперативного реагирования на различного рода изменения при перемещении роботизированной платформы. Для чего робот-помощник должен был научиться принимать решения только по части имеющейся у него информации о габаритах предметов, игнорируя незначительные детали обстановки. Именно поэтому достаточно большой объем работы пришелся на формирование датасета для обучения нейросети.

Первый шаг включал автоматическое извлечение из непрерывной видеосъемки набора кадров для обучающей выборки с соблюдением ряда требований к их содержанию и качеству. Для решения этой задачи требовалось выполнить нарезку кадров из видеопотока, провести их селекцию и обработку, включая удаление статичных кадров съемки с одинаковым сюжетом, устранение вибрационных и краевых искажений, коррекцию перспективы и т.п. [13].

Выполнение этой работы было осуществлено с помощью библиотеки алгоритмов компьютерного зрения OpenCV с открытым исходным кодом, реализованной на C/C++ [14].

Как правило, из-за особенностей объектива, а также угла съемки часть изображения по краям кадра имеет деформации, в силу чего дальнейшее использование этой части кадра нецелесообразно. Для коррекции перспективы в OpenCV применяется функция `warpPerspective`, в которую в качестве входных параметров передается исходное изображение, а также координаты прямоугольника, относительно которого необходимо выполнить трансформацию изображения с целью устранения упомянутых искажений.

Выделение интересующей нас содержательной области с отсечением лишней информации является базовым при подготовке любой обучающей выборки и положительно сказывается на финальном результате [15]. В OpenCV оно проводится с помощью функции `cvSetImageROI`, позволяющей выделить прямоугольную область, изображение в пределах которой сохраняется в базе и используется для дальнейшей работы.

Для повышения общего качества выборки проводилась её предварительная обработка с целью устранения близких по содержанию кадров. Выявление степени близости осуществлялось путем попиксельного вычитания предварительно переведенных в градацию серого изображений одного (текущего) кадра из другого (предыдущего). Соответствующая процедура в OpenCV осуществляется с использованием функции `cvtColor`. Сформированный после взаимного вычитания результирующий кадр служит основой для расчета степени соответствия изображений друг другу. Опытным путем были установлены значения критерия, при которых различия в кадре можно считать удовлетворительными.

Пример сравниваемых изображений и полученный результирующий кадр с разницей соответствия между двумя кадрами представлен на рисунках 3 и 4.



Рисунок 3: Два кадра, между которыми проводится сравнение



Рисунок 4: Результат сравнения двух кадров, где интенсивность белого тона пропорциональна степени различий

Как уже отмечалось выше, основной целью работы была разработка технологии ускоренного распознавания образов, для чего необходимо научить агента выделять существенные аспекты обстановки и игнорировать большое число незначимых деталей.

Одним из способов сжатия информационного поля с последующим выделением наиболее существенных акцентов является применение сверточных нейронных сетей [16]. Для формирования архитектуры конкретной сети в настоящей работе были использованы библиотеки с открытым исходным кодом Segnet и Tensorflow.

Для улучшения результатов обучения, заключающегося в корректном распознавании динамических объектов, их классификации и обозначении габаритов, проводится дополнительная обработка отобранных изображений различными фильтрами и применением детектора границ Санну. Все вместе это позволяет последовательно провести отсеивание шума и устранение лишних деталей из изображения, рассчитать градиенты, выделить краевые эффекты и перевести их в тонкие линии, связать полученные края в контуры. Пример дополнительной обработки отобранных кадров для формирования вспомогательного изображения представлен на рисунке 5.



Рисунок 5: Пример дополнительной обработки отобранных кадров

В итоге после первичной обработки извлеченных из видеопотока изображений полученные данные передаются в сверточную нейронную сеть для её глубокого обучения. При этом вспомогательное изображение представляет собой выделенные на кадрах типовые объекты, заполненные одним цветом (тоном).

В текущем примере для робототехнической установки, функционирующей в пределах помещений, существенными аспектами, которые необходимо различать, являются границы попадающих в поле зрения робота объектов, и их идентификация для понимания их значимости при анализе обстановки, поскольку такие объекты, как пол, стена, проход, мобильный (человек) или стационарный (препятствие) объект играют разную роль при планировании перемещений робота.

После предварительной обработки изображения с помощью детектора границ в ход вступают алгоритмы автоматической раскраски типовых зон (пол, потолок), написанных на C++ с

помощью библиотеки OpenCV. А затем, с помощью визуального анализа идет окончательная валидация размеченных зон, границ и выделение не типовых объектов, таких как люди, дверные проемы и т.д.

Таким образом, обучающая выборка для сверточной нейронной сети представляла из себя два типа изображения. Первое – исходное, и второе – сопутствующее ему сегментированное изображение, на котором объекты и границы объектов представляются однотонно в различных цветах (или в тонах в градации серого).

Пример исходного и вспомогательного изображения представлен на рисунке 6.



Рисунок 6: Исходное и вспомогательное изображения для обучения сети

Обычно сверточные нейронные сети строятся с целью детектирования тех или иных объектов. Ключевой особенностью предлагаемого метода было решение одновременно учить нейросеть распознавать не только объекты, но и границы этих самых объектов.

Исходя из свойств архитектуры и принципов обучения, работа сверточной нейронной сети обычно сводится к переходу от конкретных особенностей изображения к более абстрактным деталям. Выражаясь более простым языком, нейронная сеть упрощает те или иные объекты и элементы изображения до простых составляющих – линий, точек и т.д. Таким образом, детектирование границ является задачей, на которой подобного рода нейронные сети показывают достаточно хороший результат. При этом подобное нововведение положительно сказывается на других слоях нейронной сети, которые нацелены на распознавание объектов в кадре.

Для того, чтобы убедиться в положительном эффекте данного нововведения, были обучены две разные нейронные сети с одинаковыми параметрами. Одна была обучена распознавать различные объекты на изображении, в другую же подавался слегка измененный датасет, в котором дополнительно помимо самих объектов также были выделены границы объектов. Результат обработки изображения нейронной сетью со встроенным детектированием границ и без него представлен на рисунке 7, слева и справа соответственно.



Рисунок 7: Результат обработки изображения нейронной сетью со встроенным детектированием границ и без него.

После чего полученные кадры сравнивались друг с другом на предмет различий. Получаемое в ходе взаимного вычитания кадров изображение характеризует различия нейронных сетей в восприятии объектов и окружения.

Результат сравнения работы двух нейронных сетей представлен на рисунке 8.

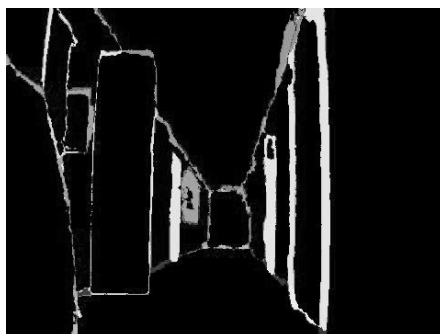


Рисунок 8: Результат сравнения работы двух нейронных сетей

Для того, чтобы вычислить насколько нейронная сеть, умеющая распознавать границы, корректно отображает действительность, полученное изображение соотносится с реальной картинкой и сопутствующим ей сегментированным изображением. В ходе чего происходит попиксельное сравнение полученного кадра и сегментированного изображения с целью выявления некорректного распознавания объектов и их границ. Таким образом, мы получаем процент соответствия полученного нейронной сетью изображения действительности.

Стоит понимать, что в связи с тем, что эксперимент проводится с нейронными сетями, преимущества и недостатки от внедрения того или иного подхода и метода можно списать на качество обучения, а также то обстоятельство, что при работе с нейронными сетями мы, по сути, имеем дело с моделью "черного ящика". Именно поэтому есть вероятность того, что на других датасетах может быть получен другой результат.

Тем не менее, на полученных в ходе решения поставленной задачи данных корректность распознавания объектов и их границ нейронной сетью в зависимости от исследуемых кадров удалось повысить в среднем на 15-20 процентов, причем в отдельных случаях увеличение точности доходило и до 30 процентов.

Получаемые по итогам обучения сети результаты работы программного комплекса представлены на рисунке 9, где по горизонтали последовательно приведены исходное изображение, вспомогательное изображение с распознанными объектами и результирующее изображение, полученное на основе выделения границ объектов обучаемым программным комплексом.

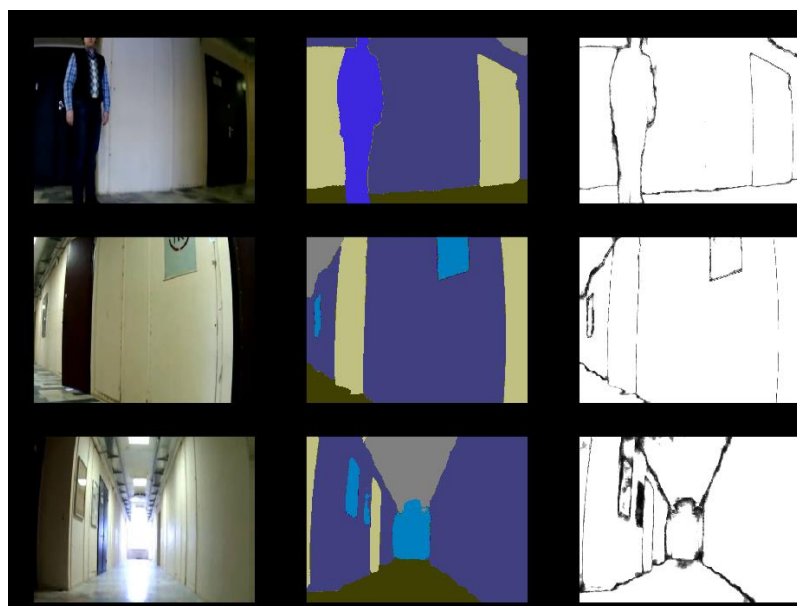


Рисунок 9: Примеры результирующих изображений с классификацией объектов и выделением их границ.

Объем данных, описывающих результирующее изображение, на несколько порядков меньше объема данных исходного, тем не менее этих данных достаточно для оперативной оценки обстановки и управления перемещением робота.

Таким образом, уменьшение объема обрабатываемой информации, позволяющее существенно ускорить процессы оценки обстановки и выработки решений на совершение рациональных действий интеллектуальным агентом в текущей ситуации, может быть достигнуто за счет выделения однородных в разрезе каких-то характеристик областей и оперирования ими как цельным объектом только на основании выявленных границ, что, по сути, эквивалентно снижению размерности обрабатываемых массивов данных.

3. Заключение

В ходе проведенного исследования была разработана методика и алгоритмы для создания обучающей выборки с последующим ее применением в рамках программного комплекса по ускоренной классификации объектов на изображениях, получаемых путем видеосъемки.

Модифицированная технология классификации и анализа обстановки с помощью сверточных нейронных сетей, описанная в настоящей работе, может быть полезна и применима не только во многих направлениях робототехники и обработки изображений, но и в ряде других задач, где важна не столько точность, сколько скорость отклика для принятия своевременных решений.

Стоит заметить, что практическое применение описанного здесь подхода связано с определенными сложностями. Так, относительно качественное выделение границ на объектах изображения на данный момент возможно лишь для ближней зоны и при достаточно высоком качестве изображения. Да и формирование такого рода обучающей выборки является весьма трудоемким и сложным процессом, нуждающимся в оптимизации.

Тем не менее, можно надеяться, что дальнейшее развитие предложенного метода позволит добиться увеличения точности обработки изображений для нахождения объектов с помощью нейронной сети, обеспечивая сохранение точности распознавания на приемлемом уровне, что весьма важно во многих ситуациях, требующих ускоренного реагирования на изменение обстановки.

4. Благодарности

Работа выполнена при финансовой поддержке РФФИ, грант № 20-31-90031.

5. Литература

- [1] В.И. Балута, А.А. Карандеев, В.П. Осипов, Функционал электронного полигона неоконфликтологии, Материалы XIII Международной конференции по прикладной математике и механике в аэрокосмической отрасли (АММАГ'2020), Алушта, 2020, с. 674. URL:http://www.npnj.ru/files/npnj2020_web.pdf.
- [2] A.A. Karandeev, V.I. Baluta, V.P. Osipov, Electronic Training Polygon for Artificial Intelligence Systems, Proceedings of the 8th International Conference on Computing for Physics and Technology (CPT2020), Nizhny Novgorod, Russia, 2020, pp. 188-192. doi:https://doi.org/10.30987/conferencearticle_5fce2771df93d0.19570965.
- [3] Д.А. Новиков. Теория управления организационными системами. 2-е изд. 2007, с. 584.
- [4] А.А. Воронин, М.В. Губко, С.П. Мишин, Д.А. Новиков. Математические модели организаций. 2008.
- [5] P. Lindes, Intelligence and Agency, Journal of Artificial General Intelligence, 11(2020). doi:[10.2478/jagi-2020-0003](https://doi.org/10.2478/jagi-2020-0003).
- [6] С. Рассел, П. Норвиг, Искусственный интеллект: современный подход, 2-е изд. пер. с англ., 2006.

- [7] A. Stocco, C. Sibert, Z. Steine-Hanson, N. Koh, J.E. Laird, C.J. Lebiere, P. Rosenbloom, Analysis of the human connectome data supports the notion of a "Common Model of Cognition" for human and human-like intelligence across domains. *NeuroImage* 235(118035), 2021, pp. 1-15.
- [8] R. Rosenbloom, A Cognitive Odyssey: From the Power Law of Practice to a General Learning Mechanism and Beyond. In *Tutorials in Quantitative Methods for Psychology*, 2006, pp. 6-14.
- [9] Г.К. Вороновский, К.В. Махотило, С.Н. Петрашев, С.А. Сергеев, Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности, 1997, с. 112.
- [10] D. Zhdanov, I. Potemin, A. Zhdanov, V. Galaktionov, Methods for the synthesis of realistic images formed by optical devices containing hologram optical elements, *Proceedings of the 8th International Conference on Computing for Physics and Technology (CPT2020)*, volume 2763 of Conference Notes, Nizhny Novgorod, Russia, 2020, pp. 30-35, ISBN 978-5-6042891-4-3. URL:http://ceur-ws.org/Vol-2763/CPT2020_paper_p-4.pdf, doi:https://doi.org/10.30987/conferencearticle_5fce2770b308b1.97036102.
- [11] B. Barladian, N. Deryabin, A. Voloboy, V. Galaktionov, L. Shapiro, High Speed Visualization in the JetOS Aviation Operating System Using Hardware Acceleration, *Proceedings of the 30th International Conference on Computer Graphics and Vision*, volume 2744 of Conference Notes, Saint Petersburg, Russia, 2020, ISSN 1613-0073, URL:<http://ceur-ws.org/Vol-2744/short3.pdf>.
- [12] P. Ramaraj, J.E. Laird. Establishing Common Ground for Learning Robots. *RSS 2018: Workshop on Models and Representations for Natural Human-Robot Communication*, Pittsburgh, PA, 2018.
- [13] D.A. Forsyth, J.Ponce, *Computer Vision: A Modern Approach*. Pearson Education. 2012.
- [14] A. Kaehler, G. Bradski, *Learning OpenCV: Computer Vision in C++ with the OpenCV Library*, 2015, pp. 575.
- [15] Nielsen, A. Michael, *Neural Networks and Deep Learning*. Determination Press, 2015.
- [16] Y. Seldin, S. Starik, M. Werman, Unsupervised Clustering of Images using their Joint Segmentation, *Proceedings of the 3rd International Workshop on Statistical and Computational Theories of Vision (SCTV)*. URL:http://www.cs.huji.ac.il/%7Ewerman/Papers/SSW_SCTV03.pdf.