ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ ДЛЯ УПРАВЛЕНИЯ МНОГОКОЛЕСНЫМ МОБИЛЬНЫМ РОБОТОМ

УДК 004.853

В.В. Демин, А.С. Кабыш, В.А. Головко,
ОИПИ НАН Беларуси, г. Минск

Аннотация

В статье рассматривается новый подход к управлению многоколесными мобильными платформами. В основе подхода лежит декомпозиция платформы на множество агентов, представляющих реальные физические колеса или блоки колесных модулей. Обучение агентов производится на основе мультиагентного обучения с подкреплением. Как подтверждение работоспособности подхода, показана серия успешных экспериментов и процесс обучения.

Введение

Эффективное управление многоколесными роботами – важная задача на производстве. Основными проблемами при построении программных систем таких роботов являются энергопотребление и планирование пути. Подсистема управления движением должна использовать эффективную модель для уменьшения энергопотребления. Этого можно добиться четырьмя способами:

– усовершенствование энергопотребления моторов. Основа таких подходов – проектирование и создание электромоторов с улучшенным коэффициентом полезного действия и увеличенным сроком службы [1, 2];

– эффективное управление движением. Классическое решение данной задачи – это расчет обратной кинематической задачи для системы управления роботом. Од-нако динамическая модель робота всегда сложнее его кинематической модели [3]. Создать более сложную модель могут помочь интеллектуальные адаптивные алгоритмы, подстраивающиеся под реального робота [4];

– эффективное планирование пути. Кратчайшая траектория на известной карте оказывается не всегда эффективной для габаритных мобильных роботов, так как их маневренность ограничена, а инерция не всегда позволяет моментально остановиться. В своих исследованиях Y. Mei и другие показали, как рассчитать энергоэффективную траекторию, используя знания об энергопотреблении управляющих воздействий [5]. S. Ogunniyi и M.S. Tsoeu продолжили данную работу, применив для расчета траектории алгоритм обучения с подкреплением;

– эффективное исследование пространства. Требуется исследовать указанную часть пространства и достигнуть целевой точки в неизвестном помещении. Роботу необходимо построить такую политику, чтобы как можно быстрее исследовать карту и достичь условленной точки [7].

В данной статье показан подход к управлению многоколесной платформой на основе многоагентного подкрепляющего обучения. Проведены эксперименты с обучением на реальном роботе. Показаны успешные испытания системы управления движением робота.

Многоколесная производственная платформа

Для решения задачи перевозки тяжелых грузов на производствах все чаще применяются автономные мобильные грузовые платформы. Одна из таких платформ – производственный грузовой робот, разработан в лаборатории университета Равенсбург-Вайнгартена [3]. Фотография робота представлена на рисунке 1 а. Основные характеристики платформы: размер платформы – 1200 см на 800 см, максимальная грузоподъемность – 500 кг при комплектации 4-мя модулями, емкость аккумуляторов – 52 Ah, независимое управление каждым модулем.

Платформа построена на базе колесных модулей [3]. Такой модуль (рисунок 1 б) состоит из двух колес, приводимых в движение двумя независимыми моторами, и имеет дифференциальную схему управления. К платформе такие модули подсоединены подшипником (рисунок 2), что позволяет им поворачиваться относительно платформы на любой угол.

а б

Рисунок 1 – Производственная грузовая мобильная платформа (а); инновационный модуль (б)

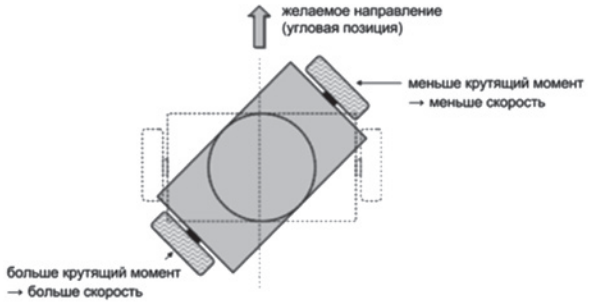

Рисунок 2 – Маневренность моторного модуля

Представленная платформа использует четыре модуля, но так же возможно собрать платформу и с большим количеством модулей.

Многоагентная система колесных модулей-агентов

Т.к. модули идентичны между собой и способ крепления к платформе одинаков, то после декомпозиции можно считать таких агентов голономными. Голономные агенты – это агенты, у которых действия и состояния совпадают. Агенты с таким свойством обладают двумя важными особенностями:

– знания одного агента можно полностью передать второму агенту, при этом второй агент становится точной копией первого;

– множество голономных агентов можно обучать взаимодействию одновременно, используя общую базу знаний.

Для получения эффекта от второго свойства будем применять архитектуру многоагентной системы с использованием виртуального лидера [8]. Это позволит накапливать базу знаний только у одного агента.

Основная идея – определение виртуального лидера, который расположен в центре формации относительно всей группы, и, соответственно, его виртуальные координаты. Состояние каждого агента будет определяться относительно виртуального лидера или виртуального центра координат. После того, как определена динамика виртуального лидера, агенты начинают свое движение. Таким образом, задача планирования пути и построения траекторий может быть реализована только для виртуального агента. Тогда как для модулей агентов будет решаться задача удержания формации во время движения. Будем называть N колесных модулей-агентов с виртуальным лидером, которые образуют многоагентную систему, платформой.

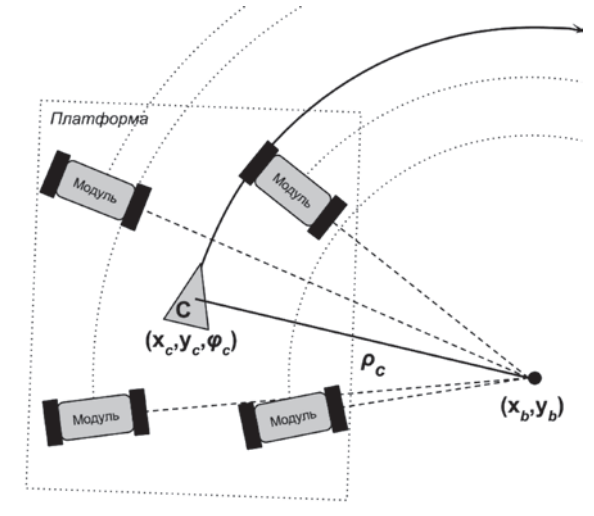

Рисунок 3 – Структура платформы с четырьмя модулями и виртуальным лидером

На рисунке 3 изображен пример такой структуры, состоящей из четырех модулей, где (x_b, y_b) – это координаты маяка, C представляет собой координаты виртуального лидера (x_c, y_c), φ_c – угол ориентации лидера относительно маяка, и ρ_c – центр разворота. Положение виртуального центра определяется как центроид площади платформы.

Обучение модулей-агентов

Процесс обучения агентов движению вокруг маяка состоит из двух этапов: (а) – обучение агента повороту на заданный угол ориентации и (б) – обучение всех агентов платформы движению по кругу за виртуальным лидером. Обучение одиночного агента повороту на угол нужной ориентации проводилось с использованием стандартного Q-learning алгоритма без следов преемственности [9].

При обучении состояние агента характеризуется парой значений $s_t = [\varphi_{err}{}^t, \omega_t]$. Множество действий $A_\omega = \{\varnothing, \omega_+, \omega_-\}$, где действие агента $a_t \in A_\omega$ – это изменение угловой скорости $\Delta\omega_t$ для момента времени t.

Система обучения сообщает положительное подкрепление, если ориентация робота ближе к целевой ($\varphi_{err}{}^t \to 0$), используя оптимальную скорость $\omega_t \to \omega_{opt}$ и штраф, если ориентация агента отклоняется от целевой, или выбранное действие не оптимально для текущего положения (агент не начал вовремя тормозить).

Значение награды определяется как:

$$r^t = R(\varphi_{err}^{t-1}, \omega^{t-1}), \quad (1)$$

где R – функция награды, которая представлена деревом принятия решений, изображенном на рисунке 4. Оптимальная скорость ω_{opt} представляет собой константное значение скорости и используется, чтобы показать, что функция ценности способна находить эффективную скорость. Для реального робота используется функция энергопотребления моторов, рассчитанная по их документации.

Структура взаимодействия агентов на основе обучения с подкреплением для решения задачи кооперативного движения изображена на рисунке 5. Модуль i, находясь в состоянии s_i^t, выбирает действие α_i^t, используя текущую стратегию выбора действий, и переходит в следующее состояние s_i^{t+1}. Виртуальный лидер получает данные об изменениях после выполнения действия, вычисляет и присваивает награду r_i^{t+1} каждому агенту, которая оценивает корректность его действий с точки зрения всей платформы.

На рисунке 6 (x_i, y_i) и (x_i^{opt}, y_i^{opt}) представляют координаты реального и целевого положения i-ого модуля со-

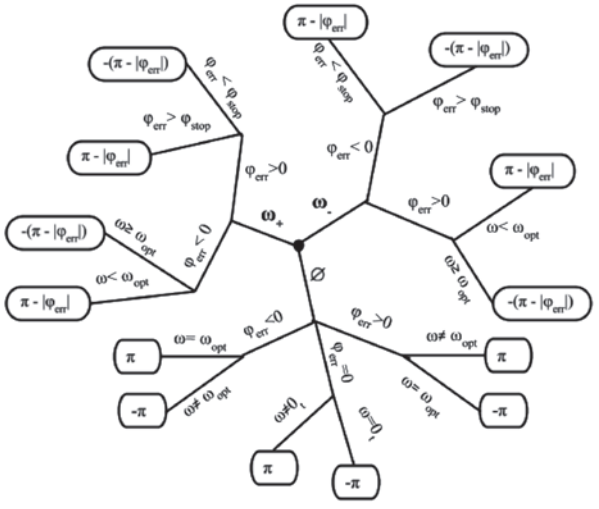

Рисунок 4 – Функция награды в виде дерева решений

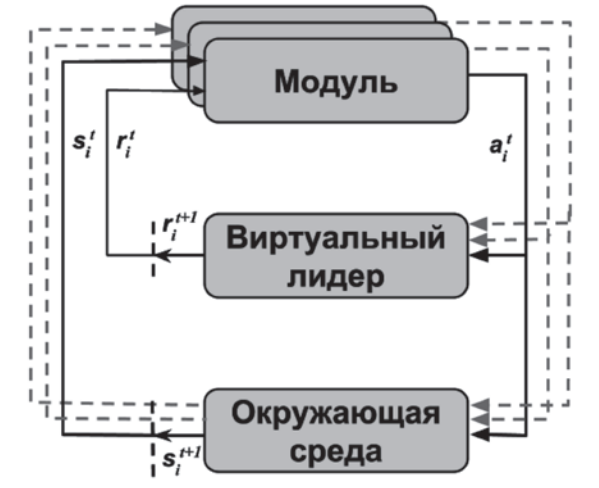

Рисунок 5 – Архитектура многоагентной системы колесных модулей на основе виртуального лидера и обучения с подкреплением

ответственно, $\bar{d}_i^{\,err}$ представляет вектор отклонения для *i*-ого модуля от правильного положения в платформе (2):

$$\bar{d}_i^{\,err} = \bar{d}_i^{\,t} - \bar{d}_i^{\,opt}, \quad \textbf{(2)}$$

где $\bar{d}_i^{\,t}$ – вектор расстояния до виртуального центра от текущего положения модуля, и $\bar{d}_i^{\,opt}$ – вектор эталонного расстояния между виртуальным центром и *i*-ым агентом, которое получено из топологии платформы.

Для обновления стратегии модулей используется аналог многоагентного *Q*-learning алгоритма [10–12], в котором виртуальный лидер начисляет награду каждому агенту как оценку состояния в платформе:

$$\Delta Q_i(s_i^t, a_i^t) = \alpha[r_i^{t+1} + \gamma \max_{a \in A(s_i^{t+1})} Q_i(s_i^{t+1}, a) - Q_i(s_i^t, a_i^t)]. \quad \textbf{(3)}$$

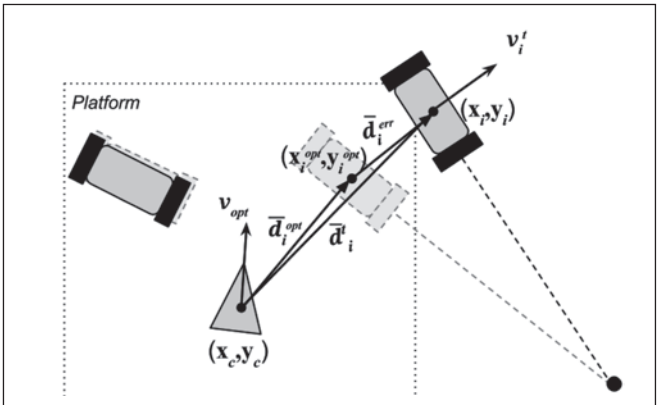

Рисунок 6 – Состояние платформы для *i*-го модуля

Результаты онлайн-обучения модуля повороту с константными скоростями

Важным аспектом обучения мультиагентной системы, предназначенной для управления реальным роботом, является онлайн-обучение системы, т.е. обучение в реальном времени. Для этих целей были проведены эксперименты с константными скоростями, чтобы сократить время обучения. Множество действий представлено в таблице 1.

Таблица 1 – Действия модуля-агента для онлайн-обучения

№	Действия робота	Value
1	Установить положительную константную скорости поворота, $+\omega$	0,8 рад/с
2	Установить отрицательную константную скорости поворота, $-\omega$	–0,8 рад/с
3	Установить скорость 0	0 рад/с

Топология *Q*-функции, которая обучалась в течение 360 эпох, показана на рисунке 7. Данная функция была протестирована на реальном роботе и решает поставленную задачу поворота модуля на небольших скоростях (для колеса до 0,2 м/с). При больших скоростях модуль не успевает остановиться в нужной точке из-за инерции, для чего используется обучение управления скоростями, показанное в следующем разделе.

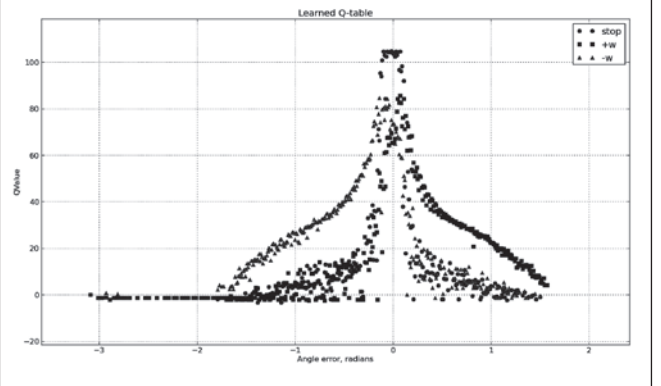

Рисунок 7 – Топология *Q*-функции после онлайн-обучения одного модуля с константными скоростями

Результаты экспериментов с производственным роботом

Для скоростей колеса свыше 0,2 м/с, как и низких скоростей, необходимо точное позиционирование модуля относительно маяка. Для минимизации погрешности агент обучается управлению скоростями. Первоначально агент обучается разгоняться до заданной системой управления скорости и поддерживать ее. Вторым этапом агент обучается уменьшать скорость вплоть до остановки, чтобы угловая ошибка была максимально приближена к нулю. Топология *Q*-функции, которая обучалась в течение 1400 эпох, показана на рисунке 8. Обучение происходило на реальном роботе.

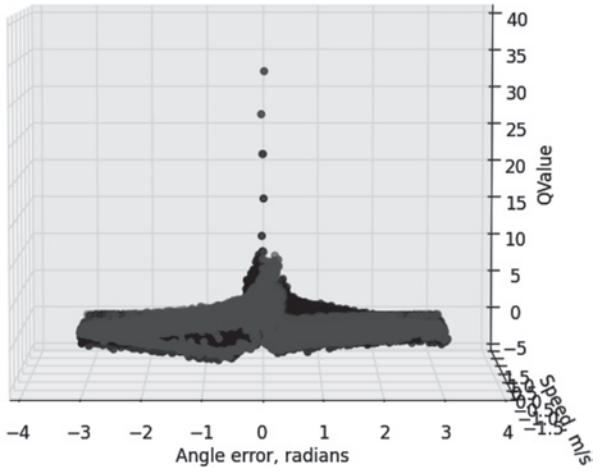

Рисунок 8 – Топология *Q*-функции после онлайн-обучения одного модуля с управлением скорости

Результат управления поворотом обученных модулей с центром разворота впереди-справа – на рисунке 9.

Знания агентов, обученных поддерживать скорость относительно виртуального лидера, были перенесены в систему управления реального робота. На рисунке 10 изображен процесс эксперимента, где на скриншотах 1-6 модули поворачиваются к маяку, изображенному на белом листке бумаги. После того, как все модули успешно выполнили задачу поворота, платформа начинает движение вокруг маяка (скриншоты 7–9, рисунок 10).

На рисунке 11 показано, что при передвижении платформы радиусы модулей не изменяются и модули платформы находятся в стабильном положении относительно платформы.

Выводы

Рассмотренный подход к применению алгоритма обучения с подкреплением в системе управления является альтернативой для расчета обратной кинематической задачи. Подход отличается адаптивностью и самонастраиваемостью к любому типу многоколесных роботов, масштабируемостью на множество движимых модулей или колес. Обучение в реальном времени позволяет вычислять лучшие политики управления для полной динамической модели робота. Ключевой особенностью алгоритма является возможность оптимизации системы по различным параметрам (энергопотребление, скорость передвижения, плавность езды). Чем больше параметров для оптимизации необходимо учесть, тем больше времени придется затратить на обучение системы, что, несомненно, является недостатком. Однако обучение необходимо производить один раз и без присутствия специалиста – система может сама вычислить все необходимые параметры. Эксперименты с реальным роботом показали возможность практического применения данного подхода.

Литература:

1. Andreas, J.C. Energy-Efficient Electric Motors, Revised and Expanded / J.C. Andreas // J.C. CRC Press. – 1992.

2. De Almeida, A.T. Energy efficiency improvements in electric motors and drives / A.T. de Almeida, P. Bertoldi, and W. Leonhard // Springer Berlin, 1997.

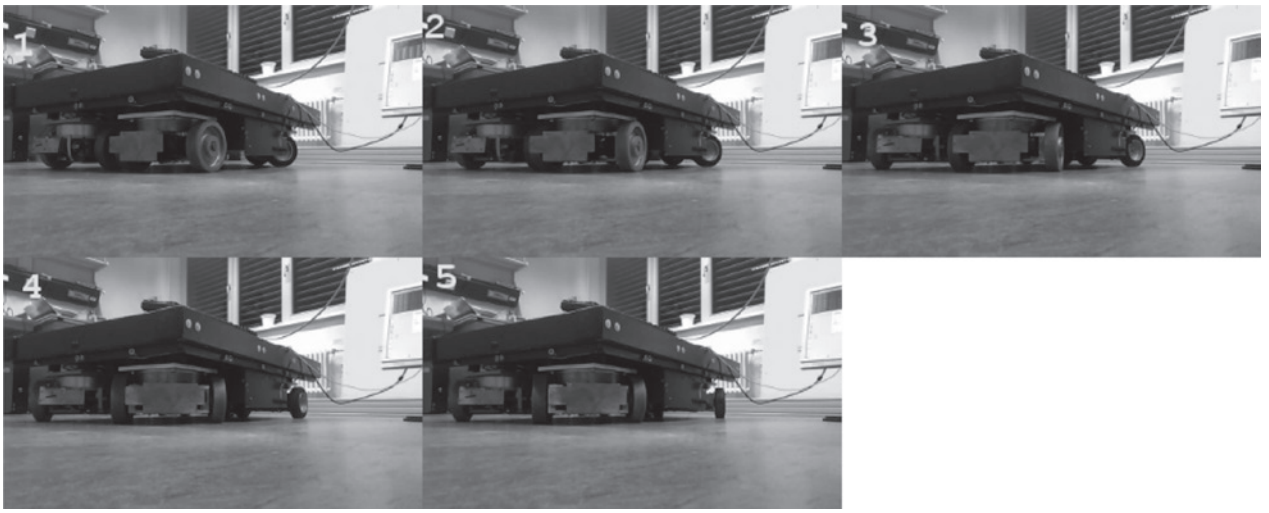

Рисунок 9 – Эксперимент поворота модулей к центру разворота впереди-справа

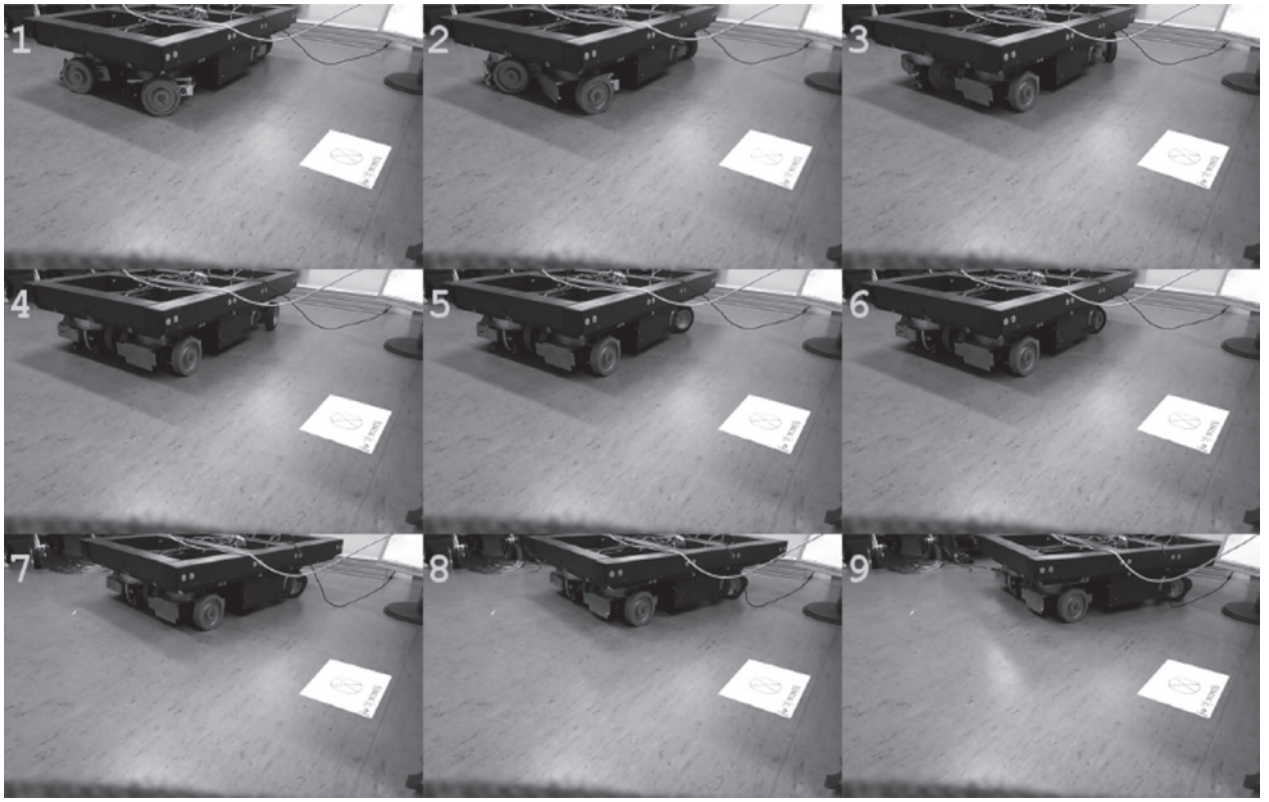

Рисунок 10 – Эксперимент поворота модулей для движения с автомобильной кинематической схемой (1–6) и движение вокруг белого маяка (7–9)

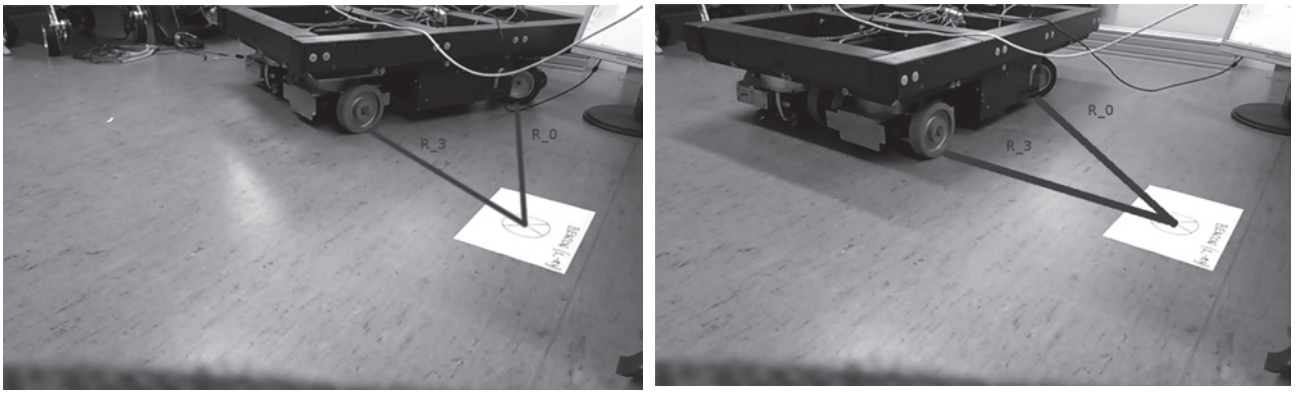

Рисунок 11 – Эксперимент подтверждения сохранности радиусов во время движения

3. Stetter, R. Development, Realization and Control of a Mobile Robot / R. Stetter, P. Ziemniak, and A. Paczynski // In Research and Education in Robotics-EUROBOT 2010, Springer, 2011:130-140.

4. Dziomin, U. A multi-agent reinforcement learning approach for the efficient control of mobile robot / U. Dziomin, A. Kabysh, V. Golovko, and R. Stetter // In Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), 2013 IEEE 7th International Conference on, 2, 2013:867-873.

5. Mei, Y. Energy-efficient motion planning for mobile robots / Y. Mei, Y.-H. Lu, Y.C. Hu, and C.G. Lee // In Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on, 5, 2004:4344-4349.

6. Ogunniyi, S. Q-learning based energy efficient path planning using weights / S. Ogunniyi, M.S. Tsoeu // In proceedings of the 24th symposium of the Pattern Recognition association of South Africa, 2013:76-82.

7. Mei, Y. Energy-efficient mobile robot exploration/ Y. Mei, Y.-H. Lu, C.G. Lee, and Y.C. Hu // In Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on, 2006:505-511.

8. Ren, W. Distributed coordination architecture for multi-robot formation control / W. Ren, N. Sorensen // Robotics and Autonomous Systems. – 2008. –Vol. 56, № 4. – P. 324–333.

9. Sutton, Richard S. Reinforcement Learning: An Introduction / Richard S. Sutton, Andrew G. Barto. – MIT Press, 1998.

10. Kabysh, A. General model for organizing interactions in multi-agent systems / A. Kabysh, V. Golovko // International Journal of Computing. –2012. – Vol. 11(3). – P. 224–233.

11. Kabysh, A. Influence Learning for Multi-Agent Systems Based on Reinforcement Learning / A. Kabysh, V. Golovko // International Journal of Computing. – Vol. 11(1). – P. 39–44.

12. Kabysh, A. Influence model and reinforcement learning for multi agent coordination / A. Kabysh, V. Golovko, K. Madani // Journal of Qafqaz University, Mathematics and Computer Science. – 2012. – № 33. – P. 58–64.

Abstract

This paper presents a novel approach of a multi-wheeled mobile robot control. The approach is based on the platform decomposition into a multi-agent system, where an agent presents a real physical wheel or a driving module. The agents learn by the multi-agent reinforcement learning algorithm to cooperate together and to provide robust steering. Experiments show that developed system provides robust control for complex dynamic robots models.

Поступила в редакцию 14.05.2014 г.

ТРЕБОВАНИЯ К НАУЧНЫМ СТАТЬЯМ, ПУБЛИКУЕМЫМ В РАЗДЕЛЕ «РЕЦЕНЗИРУЕМЫЕ СТАТЬИ»

1. Научная статья – законченное и логически цельное произведение по раскрываемой теме – должна соответствовать одному из следующих научных направлений: информационные технологии и системы, оптоэлектроника, микро- и наноэлектроника, приборостроение.

2. Объем научной статьи не должен превышать 0,35 авторского листа (14 тысяч печатных знаков, включая пробелы между словами, знаки препинания, цифры и другие), что соответствует 8 страницам текста, напечатанного через 2 интервала между строками (5,5 страницы в случае печати через 1,5 интервала).

3. Статьи в редакцию представляются в двух экземплярах на бумаге формата А4 (220015, г. Минск, пр. Пушкина, 29Б), а также в электронном виде (e-mail: sadov@bsu.by). К статье прилагаются сопроводительное письмо организации за подписью руководителя и акт экспертизы. Статья должна быть подписана всеми авторами.

Статьи принимаются в формате doc, rtf, pdf, набранные в текстовом редакторе word, включая символы латинского и греческого алфавитов вместе с индексами. Каждая иллюстрация (фотографии, рисунки, графики, таблицы и др.) должна быть представлена отдельным файлом и названа таким образом, чтобы была понятна последовательность ее размещения. Фотографии принимаются в форматах tif или jpg (300 dpi). Рисунки, графики, диаграммы принимаются в форматах tif, cdr, eps или jpg (300 dpi, текст в кривых). Таблицы принимаются в форматах doc, rtf или Excel.

4. Научные статьи должны включать следующие элементы:

аннотацию; фамилию и инициалы автора (авторов) статьи, ее название; введение; основную часть, включающую графики и другой иллюстративный материал (при их наличии); заключение; список цитированных источников; индекс УДК; аннотацию на английском языке.

5. Название статьи должно отражать основную идею выполненного исследования, быть по возможности кратким, содержать ключевые слова, позволяющие индексировать данную статью.

6. Аннотация (100–150 слов) должна ясно излагать содержание статьи и быть пригодной для опубликования в аннотациях к журналам отдельно от статьи.

В разделе «Введение» должен быть дан краткий обзор литературы по данной проблеме, указаны не решенные ранее вопросы, сформулирована и обоснована цель работы.

Основная часть статьи должна содержать описание методики, аппаратуры, объектов исследования и подробно освещать содержание исследований. Полученные результаты должны быть обсуждены с точки зрения их научной новизны и сопоставлены с соответствующими известными данными. Основная часть статьи может делиться на подразделы (с разъяснительными заголовками).

Иллюстрации, формулы, уравнения и сноски, встречающиеся в статье, должны быть пронумерованы в соответствии с порядком цитирования в тексте.

В разделе «Заключение» должны быть в сжатом виде сформулированы основные полученные результаты с указанием их новизны, преимуществ и возможностей применения.

Список цитированных источников располагается в конце текста, ссылки нумеруются согласно порядку цитирования в тексте. Порядковые номера ссылок должны быть написаны внутри квадратных скобок (например: [1], [2]).

В соответствии с рекомендациями ВАК Республики Беларусь от 29.12.2007г. №29/13/15 научные статьи аспирантов последнего года обучения публикуются вне очереди при условии их полного соответствия требованиям, предъявляемым к рецензируемым научным публикациям.