

УДК 004.896, 004.5

## ПИЛОТНЫЕ ИССЛЕДОВАНИЯ ПО УПРАВЛЕНИЮ РОБОМОБИЛЕМ AVRORA UNIOR С ПОМОЩЬЮ ПРОИЗВОЛЬНЫХ ЖЕСТОВ

Т.Г Цой (*tt@it.kfu.ru*)

Н.А. Никифоров (*nanikiforov1337@gmail.com*)

О.С Прокопчук (*oks.p@it.kfu.ru*)

Р.Н. Сафин (*safin.ramil@it.kfu.ru*)

Е.А Магид (*magid@it.kfu.ru*)

Лаборатория Интеллектуальных Робототехнических Систем,  
Институт информационных технологий и интеллектуальных  
систем, Казанский федеральный университет, Казань

Я. Бай (*yangbai@fc.ritsumei.ac.jp*)

М.М. Свинин (*svinin@fc.ritsumei.ac.jp*)

Кафедра информационных наук и инженерии, Колледж  
информационных наук и инженерии, Университет  
Рицумейкан, Кусацу, Япония

**Аннотация.** Невербальное общение при помощи жестов является важным способом коммуникации, который используется при взаимодействии с интеллектуальными устройствами. В статье представлена концепция управления движением робомобиля Avroga Unior с помощью произвольных жестов. Группа из 17 добровольцев получила список из 18 команд управления и независимо друг от друга использовала его для создания произвольных индивидуальных жестов управления. Часть полученного набора жестов использовалась с платформой Teachable machine для оценки возможности построения модели и ее обучения. Полученная модель продемонстрировала высокую скорость распознавания. Были также апробированы методы SURF и FLANN для прямого сопоставления жестов и подход к сопоставлению на основе скелетонизации.<sup>1</sup>

---

<sup>1</sup> Работа выполнена при поддержке Российского фонда фундаментальных исследований (РФФИ), проект ID 19-58-70002. Шестой и седьмой авторы выражают признательность за поддержку Японскому агентству по науке и технологиям, Программа стратегических международных совместных исследований JST, проект № 18065977.

**Ключевые слова:** управление на основе жестов, управление роботом, машинное обучение, датасет, мобильный робот, роботомобиль Avrora Unior.

## Введение

Системы компьютерного зрения имеют широкое применение в задачах распознавания автомобильных номеров [Qadri et al., 2009], меток [He et al., 2009], животных [Nguyen et al., 2017], человеческого лица [Ray et al., 2015] и персональных гаджетов [Sutoyo et al., 2016]. В поисково-спасательных операциях методы компьютерного зрения позволяют дистанционно управлять транспортными средствами [Al-Kaff et al., 2017]. Список возможностей прикладного использования систем технического зрения (СТЗ) постоянно пополняется по мере возникновения задач, для решение которых не хватает ограниченных человеческих возможностей точного и быстрого обнаружения и распознавания.

Распознавание жестов часто используется в качестве важного канала коммуникации не только при взаимодействии человека с человеком, но и при взаимодействии человека с роботом [Shirwalkar et al., 2013] или коммуникации человека с другими интеллектуальными устройствами [Rashid et al., 2016]. В [Rahman et al., 2009] авторы продемонстрировали систему интерфейса для управления электроприборами в среде умного дома при помощи жестов рук. В [Muñoz-Salinas et al., 2008] камера глубины извлекала силуэт руки, а полученные изображения были распознаны с помощью обученного алгоритма random forest [Pal, 2005]. В [Rautaray, 2012] система распознавания жестов генерировала соответствующую команду, которая позволяла сделать выбор и управлять мышью.



Рис. 1. Робот Avrora Unior. Красным прямоугольником отмечен сенсор Microsoft Kinect.

В данной статье рассматривается взаимодействие человека и робота с использованием жестов [Galín et al., 2019]. Цель долгосрочного проекта –

разработка управления роботомобилем *Avrora Unior* (рис. 1) с помощью пользовательских жестов, которые не predeterminedены заранее [Malov et al., 2019]. Для разработки концепции управления роботомобилем жестами были изучены возможности передвижения робота и создан подробный список основных команд управления. Мы попросили группу студентов придумать свои собственные жесты, которые, по их мнению, соответствовали бы каждой команде из списка. Для распознавания жестов без построения скелета пользователя были апробированы подходы SURF [Bay et al., 2008] и FLANN [Goel et al., 2006].

## 1 Концепция управления жестами

Использование жестов для взаимодействия с роботами и интеллектуальными устройствами исследовалось во многих работах. Жесты позволяют управлять промышленными роботами с помощью удобного для пользователя интерфейса на основе жестов [Tang et al., 2018], технологии Leap Motion [Chen et al., 2015] или Microsoft Kinect Controller [Mikadlicki et al., 2015]. Даже недорогая USB-камера может успешно распознавать и отслеживать движения руки пользователя и позволяет контролировать простые действия [Grif et al., 2016]. Файо и др. продемонстрировали взаимодействие с гуманоидным роботом-помощником *NAO* с помощью статичных жестов рук [Phyo et al., 2019]. Гао и др. представили интеллектуальное управление инвалидными колясками для пожилых людей и инвалидов на основе жестов, которое было успешно протестировано в условиях закрытого помещения [Gao et al., 2017]. Жанг и др. продемонстрировали управление реальным беспилотным автомобилем на основе жестов с помощью сенсора Kinect-V2, который использует распознавание позы верхней части тела для 13 суставов [Zhang, 2012].

Используемый роботомобиль *Avrora Unior* (рис. 1) оснащен сенсором Microsoft Kinect. Важной характеристикой сенсора является диапазон, в котором гарантируются правильные значения; так, для Microsoft Kinect максимальное расстояние между объектом и датчиком ограничено 3,5 метрами [Han et al., 2013]. Это ограничение следует учитывать при разработке управляющих жестов и построении набора данных для машинного обучения [Safin et al., 2018].

Робот *Avrora Unior* передвигается автономно [Magid et al., 2017] или управляется в дистанционном режиме с помощью специального пульта, удерживаемого одной рукой. Основная цель введения дополнительного управления жестами заключается в обеспечении удобного процесса тестирования новых алгоритмов движения и взаимодействия без помощи пульта. Пользователь должен подойти к роботу на расстоянии 3–4 метров и, находясь перед ним полностью в поле зрения сенсора Microsoft Kinect,

показать определенный жест управления, который запускает выполнение соответствующей команды. Для упрощения распознавание были предложены статичные жесты управления.

Для обеспечения максимальной безопасности полноразмерных автономных транспортных средств обычно все жесты управления предопределяются заранее, а затем система управления роботом обучается распознавать эти жесты с использованием исчерпывающего набора примеров [Lavrenov et al., 2017]. Однако такой подход требует от оператора тщательного изучения жестов и постоянной концентрации, чтобы использовать подходящий. В нашем случае, мы хотели бы позволить неподготовленному пользователю управлять роботом интуитивно, что подразумевает определение роботом намерений пользователя в режиме реального времени.

На первом этапе проекта был создан список команд дистанционного управления. При отборе учитывалось несколько факторов: удобство использования команды, важность команды и возможность ее использования при небольшом расстоянии между человеком-оператором и роботом. Базовый набор содержал только движения вперед и назад, повороты влево и вправо. Затем в набор было добавлено несколько более сложных команд, например, поворот на 180 градусов, приближение к неподвижному человеку и автоматическая параллельная парковка [Imameev et al., 2020]. Набор команд управления представлен в таблице 1.



Рис. 2. Предлагаемые жесты шести пользователей для команды «Полная остановка».



Рис. 3. Примеры неподходящих жестов из-за их высоты, выходящей за пределы поля зрения камеры (слева и в центре) и неудобной позы (справа).

## 2 Набор данных

Для сбора данных необходимо было записать 30-40 секундное видео со статичным жестом для каждой команды, сохраняя расстояние от камеры до пользователя в пределах 3-4 метров, что соответствует возможностям сенсора Microsoft Kinect. Группу из 17 человек, студентов и сотрудников Лаборатории интеллектуальных робототехнических систем, независимо друг от друга попросили предложить жест для каждой из 18 команд управления. Никто из участников не знал, какие жесты выбрали другие участники. Это было сделано для того, чтобы проанализировать и выбрать наиболее подходящий жест для конкретной команды на основе полученного разнообразия жестов и их статистического распределения. Кадры были извлечены из видео с помощью программного обеспечения Matlab. В таблице 1 представлено количество кадров, которые были извлечены из собранного набора из 306 (17 человек, по 18 жестов каждый) коротких видеозаписей для каждого из жестов.

Таблица 1. Список команд управления и размер набора данных (кадров) для каждой команды.

№	Команда	Комментарий	Кадры
1	Движение вперед	Заданная постоянная скорость	30755
2	Движение назад		29011
3	Поворот колес вправо	Колеса поворачиваются при статичном состоянии, угол поворота постепенно увеличивается	30450
4	Поворот колес влево		28990
5	Увеличение скорости	Скорость постепенно увеличивается на заданное значение	29048

6	Уменьшение скорости	Скорость постепенно снижается на заданное значение	28829
7	Полная остановка	Экстренное торможение	30233
8	Включить режим игнорирования пользователя	Игнорировать команды, кроме команды полной остановки и команды отключения этого режима	30615
9	Выключить режим игнорирования пользователя	Отключить режим игнорирования	29427
10	Движение вперед, поворачивая колеса вправо	С заданной постоянной скоростью и заданным углом поворота колес	28781
11	Движение вперед, поворачивая колеса влево		30021
12	Движение назад, поворачивая колеса вправо		26581
13	Движение назад, поворачивая колеса влево		26939
14	Автоматический поворот на 180 градусов		30941
15	Автоматический поворот на 90 градусов вправо		29691
16	Автоматический поворот на 90 градусов влево		29892
17	Движение к пользователю		31205
18	Автоматическая параллельная парковка	Единственная сложная команда в наборе, которая подразумевает поиск ближайшего места парковки [Imameev et al., 2021] и дальнейшую парковку.	31847

Сходство между жестами разных пользователей было небольшим. Только одна команда, «Полная остановка», во многих случаях имела ограниченное разнообразие паттернов и существенное сходство жестов внутри паттерна (рис. 2).

### 3 Пилотные исследования

#### 3.1 Оптимальный выбор жестов

Выбор оптимальных жестов для каждой команды основывался на трех критериях: размер (ограничивающая область для всего тела пользователя при демонстрации команды), простота использования и отсутствие

перекрытия скелета пользователя звеньями верхних конечностей. Важно, чтобы ограничивающая область для жеста не превышала установленного порога, который зависит от расстояния (в данном исследовании - 3-4 м от пользователя до робота) и роста пользователя, иначе жест может оказаться вне поля зрения сенсора Microsoft Kinect, расположенного на высоте 0,35 м от земли [Shabalina et al., 2019]. Например, на рисунке 3 представлены неподходящие жесты, которые окажутся вне поля зрения камеры.

Некоторые жесты могут быть трудновыполнимыми из-за физических ограничений человеческого скелета и подвижности суставов. Это означает, что такие жесты будут редко использоваться обычным пользователем. На рисунке 3 (справа) также представлен пример жеста, который может быть трудно повторить, поскольку эта поза неудобна для обычного человека.

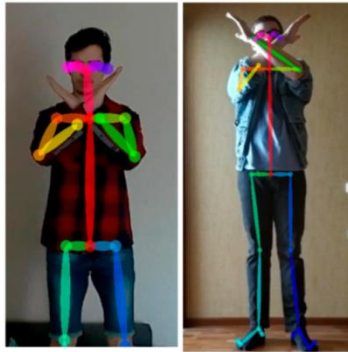


Рис. 4. Примеры неподходящих жестов из-за перекрытия звеньев скелета.

Для анализа позы пользователя и извлечения управляющего сигнала был выбран метод извлечения базового скелета. Для построения базового скелета использовалась система OpenPose [Cao et al., 2019]. Были исключены жесты, которые могут быть неправильно распознаны из-за перекрытия скелета звеньями верхних конечностей (рис. 4).

### 3.2 Оптимальный выбор жестов

Чтобы проверить, позволит ли собранный набор изображений в дальнейшем создать полномасштабную модель для машинного обучения, была использована платформа Teachable machine [Carney, et al., 2020], которая рисует скелет для набора данных изображений и позволяет построить и обучить новую модель в соответствии с выбранными параметрами обучения. На начальном этапе для каждой команды использовались только 50 изображений от 5 участников, чтобы обучить очень простую модель различать все 18 команд управления (таблица 1). Тем

не менее, даже для такого небольшого набора данных, обученная модель продемонстрировала хорошие результаты (рис. 5), которые будут улучшаться в процессе обучения на основе набора данных.

Для того чтобы проверить, возможно ли определить правильно отображаемые жесты без построения скелета и может ли использование скелета улучшить точность распознавания, мы использовали методы Speeded-Up Robust Features (SURF [Bay et al., 2008]) и Fast Library for Approximate Nearest Neighbors (FLANN [Goel et al., 2006]). SURF использовался для поиска особых точек изображений (кадров из видеозаписей). Далее, для получения быстрого и эффективного сопоставления, проводилось сравнение с использованием FLANN. На рисунке 6 показан пример сопоставления особых точек с помощью двух подходов: прямого сопоставления (верхний ряд рис.6; только 67 совпадений были успешными) и с использованием базового скелета (центральный ряд рис.6; 103 совпадения были успешными). Подход на основе скелета обеспечил хороший уровень совпадения (нижний ряд рис.6; 91 совпадение было успешным); однако в этом примере две позы объективно схожи, поэтому высокий уровень совпадения был закономерным.

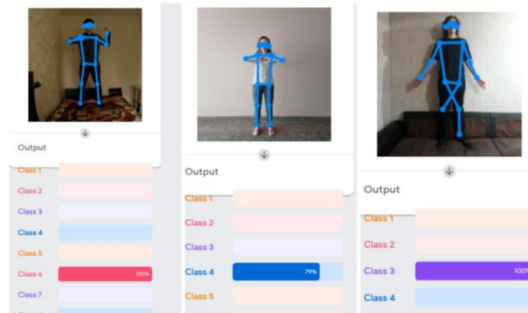


Рис. 5. Пример результатов обучения модели с использованием сервиса платформы Teachable machine: 100% успеха для класса 6, 79% для класса 4 и 100% для класса 3. Обратите внимание на ошибку в извлечении скелета для изображения справа.

Сопоставление некоторых других изображений показало, что использование метода на основе скелета не всегда может приводить к количественному улучшению по сравнению с методом прямого сопоставления. Например, рисунок 7 демонстрирует частный случай, когда прямое сопоставление поз без скелета количественно превосходит сопоставление на основе скелета (106 совпадений против 101 совпадения соответственно), но детальное рассмотрение полученных совпадений показывает, что в обоих случаях подавляющее большинство совпадений



было неверным, в то время как ситуация с правильными совпадениями была немного лучше для подхода, основанного на скелете. Те же проблемы с неверным совпадением применимы и к рисунку 6, где основанный на скелете подход (рис. 6, центральный ряд) продемонстрировал значительно большее количество правильно сопряженных особых точек по сравнению с прямым подходом. Результаты экспериментов показали, что использование методов SURF и FLANN для сопоставления, как с прямым подходом, так и с подходом на основе скелета, не обеспечивает приемлемого уровня совпадения для двух схожих поз человека. Поэтому создание нового классификатора и его соответствующее обучение становится необходимым.

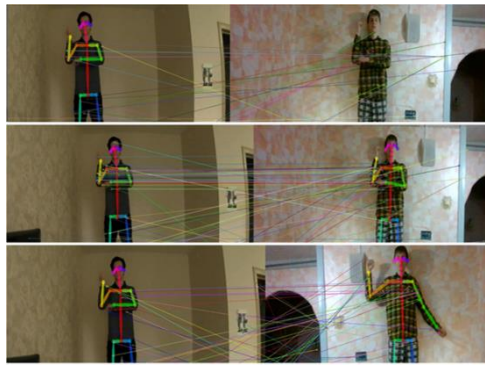


Рис. 6. Прямое сопоставление жестов без использования скелета (верхний ряд, 67 совпадений), на основе скелета (центральный ряд, 103 совпадения) и на основе скелета немного отличающегося жеста (нижний ряд, 91 совпадение).



Рис. 7. Прямое сопоставление поз без использования скелета (106 совпадений) и сопоставление на основе скелета (101 совпадение).

## Заключение

В данной статье представлена концепция управления движением робомобиля Avroga Unior с помощью жестов. Список из 18 команд управления состоял из базовых и составных команд. Группа из 17 добровольцев использовала список команд для самостоятельного создания индивидуальных жестов управления. Небольшая часть полученного набора данных жестов (менее 0,2%) использовалась с платформой Teachable machine, чтобы предварительно оценить возможность построения полномасштабной модели и обучить ее соответствующим образом. Полученная модель продемонстрировала приемлемую скорость распознавания. Была проведена апробация возможности пользования методов SURF и FLANN для применения подхода прямого сопоставления и подхода на основе скелета, но они продемонстрировали недостаточное качество результатов. На основе эмпирических исследований был сделан вывод, что собранный набор данных позволит построить хорошую модель, которую можно обучить успешно различать жесты управления движением.

## Список литературы

- [Al-Kaff et al., 2017] Al-Kaff A., Moreno F.M., de la Escalera A., Armingol J.M. Intelligent Vehicle for Search, Rescue and Transportation Purposes // IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR), China. 2017.
- [Bay et al., 2008] Bay H., Ess A., Tuytelaars T., Van Gool L. Speeded-Up Robust Features (SURF). // Computer Vision and Image Understanding. 2008. № 110(3).
- [Cao et al., 2019] Cao Z., Hidalgo G., Simon T., Wei S.E., Sheikh Y. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. // IEEE transactions on pattern analysis and machine intelligence. 2019. № 43(1).
- [Carney, et al., 2020] Carney M., Webster B., Alvarado I., Phillips K., Howell N., Griffith J., Jongejan J., Pitaru A., Chen A. Teachable Machine: Approachable Web-Based Tool for Exploring Machine Learning Classification // Conference on Human Factors in Computing Systems (CHI), USA. 2020.
- [Chen et al., 2015] Chen S., Ma H., Yang C., Fu M. Hand Gesture Based Robot Control System Using Leap Motion // Proc. 8th International Conference on Intelligent Robotics and Applications (ICIRA), UK. 2015.
- [Galın et al., 2019] Galın R., Meshcheryakov R. Review on Human–Robot Interaction During Collaboration in a Shared Workspace // Proc. 4th International Conference on Interactive Collaborative Robotics (ICR), Turkey. 2019.
- [Gao et al., 2017] Gao X., Shi L., Wang Q. The Design of Robotic Wheelchair Control System Based on Hand Gesture Control for the Disabled // Proc. 1st International Conference on Robotics and Automation Sciences (ICRAS), China. 2017.
- [Goel et al., 2006] Goel A., Saxena S.C., Bhanot S. Modified Functional Link Artificial Neural Network. // International Journal of Electrical and Computer Engineering. 2006. № 1(1).

- [Grif et al., 2016] Grif H.S., Farcas C.C. Mouse cursor control system based on hand gesture. // *Procedia Technology*. 2016. № 22.
- [Han et al., 2013] Han J., Shao L., Xu D., Shotton J. Enhanced Computer Vision with Microsoft Kinect Sensor: A Review. // *IEEE transactions on cybernetics*. 2013. № 43(5).
- [He et al., 2009] He L., Chao Y., Suzuki K., Wu K. Fast connected-component labeling. // *Pattern recognition*. 2009. № 42(9).
- [Imameev et al., 2020] Imameev D., Shabalina K., Sagitov A., Su K.L., Magid E. Modelling Autonomous Parallel Parking Procedure for Car-like Robot Avraa in Gazebo Simulator // *Proc. The 2020 International Conference on Artificial Life and Robotics (ICAROB2020)*, Japan. 2020.
- [Imameev et al., 2021] Imameev D., Zakiev A., Tsoy T., Bai Y., Svinin M., Magid E. Lidar-based parking spot search algorithm // *Proc. 13th International Conference on Machine Vision (ICMV)*, Italy. 2021.
- [Lavrenov et al., 2017] Lavrenov R., Zakiev A. Tool for 3D Gazebo map construction from arbitrary images and laser scans // *Proc. 10th International Conference on Developments in eSystems Engineering (DeSE)*, France. 2017.
- [Magid et al., 2017] Magid E., Lavrenov R., Khasianov A. Modified spline-based path planning for autonomous ground vehicle // *Proc. 14th International Conference on Informatics in Control, Automation and Robotics ICINCO*, Spain. 2017.
- [Malov et al., 2019] Malov D., Edemskii A., Saveliev A. Architecture of Proactive Localization Service for Cyber-Physical System's Users // *Proc. 4th International Conference on Interactive Collaborative Robotics (ICR)*, Turkey. 2019.
- [Mikadlicki et al., 2015] Mikadlicki K., Pajor M. Real-time gesture control of a CNC machine tool with the use Microsoft Kinect sensor. // *International Journal of Scientific & Engineering Research*. 2015. № 6(9).
- [Muñoz-Salinas et al., 2008] Muñoz-Salinas R., Medina-Carnicer R., Madrid-Cuevas F.J., Carmona-Poyato A. Depth silhouettes for gesture recognition. // *Pattern Recognition Letters*. 2008. № 29(3).
- [Nguyen et al., 2017] Nguyen H., Maclagan S.J., Nguyen T.D., Nguyen T., Flemons P., Andrews K., Ritchie E.G., Phung D. Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring // *Proc. 4th international conference on data science and advanced Analytics*, Japan. 2017.
- [Pal, 2005] Pal M. Random forest classifier for remote sensing classification. // *International journal of remote sensing*. 2005. № 26(1).
- [Phyo et al., 2019] Phyo A.S., Fukuda H., Lam A., Kobayashi Y., Kuno Y. A human-robot interaction system based on calling hand gestures // *Proc. 15th International Conference on Intelligent Computing (ICIC)*, China. 2019.
- [Qadri et al., 2009] Qadri M.T., Asif M. Automatic Number Plate Recognition System for Vehicle Identification Using Optical Character Recognition // *Proc. International Conference on Education Technology and Computer (ICETC)*, Singapore. 2009.
- [Rahman et al., 2009] Rahman A.M., Hossain M.A., Parra J., El Saddik A. Motion-path based gesture interaction with smart home services // *Proc. 17th ACM International conference on Multimedia*, China. 2009.
- [Rashid et al., 2016] Rashid M., Han X. Gesture control of ZigBee connected smart home Internet of Things // *Proc. 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, Bangladesh. 2016.

- [**Rautaray, 2012**] Rautaray S.S. Real Time Hand Gesture Recognition System for Dynamic Applications. // International Journal of UbiComp. 2012. № 3(1).
- [**Ray et al., 2015**] Ray S., Das S., Sen A. An intelligent vision system for monitoring security and surveillance of ATM // Proc. 2015 Annual IEEE India Conference (INDICON), India. 2015.
- [**Safin et al., 2018**] Safin R., Lavrenov R., Tsoy T., Svinin M., Magid E. Real-time video server implementation for a mobile robot // Proc. 11th International Conference on Developments in eSystems Engineering (DeSE), UK. 2018.
- [**Shabalina et al., 2019**] Shabalina K., Sagitov A., Su K.L., Hsia K.H., Magid E. Avrora unior car-like robot in gazebo environment // Proc. The 2019 International Conference on Artificial Life and Robotics (ICAROB2019), Japan. 2019.
- [**Shirwalkar et al., 2013**] Shirwalkar S., Singh A., Sharma K., Singh N. Telemanipulation of an industrial robotic arm using gesture recognition with Kinect // Proc. 2013 International Conference on Control, Automation, Robotics and Embedded Systems (CARE), India. 2013.
- [**Sutoyo et al., 2016**] Sutoyo R., Harefa J., Chowanda A. Unlock Screen Application Design Using Face Expression on Android Smartphone // Proc. 7th International Conference on Mechanical, Industrial, and Manufacturing Technologies, South Africa. 2016.
- [**Tang et al., 2018**] Tang G. and Webb P. The Design and Evaluation of an Ergonomic Contactless Gesture Control System for Industrial Robots. // Journal of Robotics. 2018.
- [**Zhang et al., 2020**] Zhang B., Yang M., Yuan W., Wang C., Wang B. A Novel System for Guiding Unmanned Vehicles Based on Human Gesture Recognition // Proc. International Conference on Real-time Computing and Robotics (RCAR), Japan. 2020.
- [**Zhang, 2012**] Zhang Z. Microsoft Kinect sensor and its effect. // IEEE multimedia. 2012. № 19(2).