НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
май–июнь 2026 Том 26 № 3 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
May–June 2026 Vol. 26 No 3 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2026-26-3-532-543
УДК 004.056

Метод выявления вредоносных роботов в задаче коллективного восприятия окружающей среды

Игорь Алексеевич Зикратов[1]✉, Татьяна Викторовна Зикратова[2]

[1] Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, Санкт-Петербург, 193232, Российская Федерация
[2] Военно-морской политехнический институт ВУНЦ ВМФ «Военно-морская академия имени Адмирала флота Советского Союза Н.Г. Кузнецова», Санкт-Петербург, 197342, Российская Федерация

[1] igzikratov@yandex.ru✉, https://orcid.org/0000-0001-9054-800X
[2] ztv64@yandex.ru, https://orcid.org/0000-0001-8365-658X

Аннотация

Введение. Многовариантное коллективное принятие решений в мультиагентных робототехнических системах является сложной задачей роевого интеллекта. Особенности архитектуры гомогенных групп роботов с децентрализованным управлением и ограниченные способности отдельных агентов предоставляют возможность злоумышленникам для внедрения и использования в составе роя вредоносных роботов. Вредоносные роботы применяют различные стратегии поведения и создают условия для принятия ошибочного решения в процессе достижения консенсуса. Предлагается метод, позволяющий выявить вредоносных роботов в процессе выполнения роем задачи картографирования исходной сцены. **Метод.** Предложенный подход решает задачу выявления в составе роя вредоносных роботов, независимо от используемой ими стратегии поведения. Метод основан на гипотезе, что статистические характеристики локальных карт, полученные исправными роботами, соответствуют статистическим характеристикам исходной сцены, и не соответствуют характеристикам локальных карт вредоносных роботов. В качестве атрибутов для распознавания предложены частотные гистограммы атрибутов исследуемого объекта, которые формируются при анализе локальных карт. Решение задачи распознавания осуществляется наивным байесовским классификатором. Такой подход позволяет обеспечить высокие показатели качества распознавания за счет выявления статистически значимой разницы гистограмм локальных карт исправных и вредоносных роботов. **Основные результаты.** Проведена оценка показателей качества разрабатываемого алгоритма для различных типов сцены. Выполнена серия испытаний, в которых для единых исходных данных оценивалась вероятность правильного определения вредоносных роботов, использующих различные стратегии поведения. Показано, что при предварительном обучении байесовского классификатора на сцене с идентичными статистическими характеристиками ошибка второго рода снижается до минимальных значений. Определение ошибки второго рода критически важно для исключения ситуации, когда к корректирующему обсуждению решения могут быть допущены вредоносные агенты. **Обсуждение.** Предложенный алгоритм отличается высокой степенью абстракции, что позволяет рассматривать его использование в широком круге задач коллективного восприятия окружающей среды при наличии преднамеренных вредоносных информационных воздействий.

Ключевые слова

групповая робототехника, коллективное восприятие, вредоносные роботы, безопасность мультиагентных систем, распознавание образов

Ссылка для цитирования: Зикратов И.А., Зикратова Т.В. Метод выявления вредоносных роботов в задаче коллективного восприятия окружающей среды // Научно-технический вестник информационных технологий, механики и оптики. 2026. Т. 26, № 3. С. 532–543. doi: 10.17586/2226-1494-2026-26-3-532-543

Method for detecting malicious robots in the collective perception of the environment

Igor A. Zikratov[1]✉, Tatiana V. Zikratova[2]

[1] The Bonch-Bruevich Saint Petersburg State University of Telecommunications (SPbSUT), Saint Petersburg, 193232, Russian Federation

[2] VUNTS of the Navy "Naval Academy", Saint Petersburg, 197342, Russian Federation

[1] igzikratov@yandex.ru✉, https://orcid.org/0000-0001-9054-800X

[2] ztv64@yandex.ru, https://orcid.org/0000-0001-8365-658X

Abstract

Multi-variant collective decision-making in multi-agent robotic systems is a complex problem of swarm intelligence. The architectural features of homogeneous robot groups with decentralized control and the limited capabilities of individual agents provide an opportunity for attackers to introduce and use malicious robots within the swarm. Malicious robots employ various behavioral strategies and create conditions for erroneous decision-making during consensus formation. A method is proposed to detect malicious robots while the swarm performs the task of mapping the original scene. The proposed approach solves the problem of detecting malicious robots within a swarm, regardless of the behavioral strategy they use. The method is based on the hypothesis that the statistical characteristics of local maps obtained by benign robots correspond to the statistical characteristics of the original scene and do not match the characteristics of local maps from malicious robots. Frequency histograms of attributes of the examined object, generated during the analysis of local maps, are proposed as recognition features. The recognition problem is solved using a naive Bayes classifier. This approach ensures high recognition quality by identifying statistically significant differences in the histograms of local maps from benign and malicious robots. The performance metrics of the developed algorithm are evaluated for various scene types. A series of experiments is conducted in which the probability of correctly identifying malicious robots using different behavioral strategies is assessed for the same initial data. It is shown that when the Bayesian classifier is pre-trained on a scene with identical statistical characteristics, the Type II error is reduced to minimal values. Identifying the Type II error is critically important to prevent malicious agents from being admitted to corrective decision-making discussions. The proposed algorithm features a high degree of abstraction, allowing it to be considered for use in a wide range of collective perception tasks involving deliberate malicious information attacks.

Keywords

swarm robotics, collective perception, malicious robots, multi-agent system security, pattern recognition

For citation: Zikratov I.A., Zikratova T.V. Method for detecting malicious robots in the collective perception of the environment. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2026, vol. 26, no. 3, pp. 532–543 (in Russian). doi: 10.17586/2226-1494-2026-26-3-532-543

Введение

Распределенные киберфизические системы, реализующие парадигмы «умной пыли» [1], состоят из большого количества относительно примитивных и плохо информированных компонентов, что обычно имеет место для гомогенных групп (роев) роботов. Ограничения отдельных членов роя роботов (агентов) можно преодолеть путем коллективной агрегации и обработки информации, циркулирующей в робототехнической системе, на основе которой принимается коллективное решение в отношении целевой функции [2, 3].

Для оценки эффективности различных методов коллективного восприятия принято использовать стандартные сценарии, на которых можно оценить достоинства того или иного метода организации роевого интеллекта. Высокая степень абстрагирования и общность методов коллективного восприятия позволяет проектировщику осмысленно использовать ее высокоуровневую логику управления в различных сценариях проблем, и, в результате, сосредоточиться только на реализации специфичных для предметной области низкоуровневых процедур управления [4]. Примером такого сценария является выработка дискретной коллективной оценки изучаемой среды, в процессе которой роботы перемещаются по сцене, состоящей из плиток двух или более цветов, и находясь в информационном обмене друг с другом пытаются выявить преобладающий цвет.

Известен ряд методов решения таких задач роем роботов, например Direct Modulation of Majority-based Decisions (DMMD) [5], Direct Modulation of Voter-based Decisions (DMVD) [6], и Direct Comparison [4]. Указанная группа методов представляет гибридный агентно-ориентированный подход, сочетающий в себе достоинства роевых метаэвристических алгоритмов, и методов, основанных на динамике общественного мнения [7]. На практике эти методы находят применение в мультиагентных робототехнических системах (МРТС) для решения задач координации групп роботов для разведки местности или мониторинга окружающей среды, коллективного принятия решений в ситуациях обнаружения объектов или препятствий, картографирования, совместной навигации и оптимизации маршрутов группы мобильных роботов.

Постановка задачи

В общем случае сценарии коллективного восприятия в задаче мониторинга сцены можно описать следующим алгоритмом. После запуска итерационного цикла j-й робот $r_j \in R$, где R — множество роботов группы, последовательно обходит плитки сцены. При достижении количества итераций j-го робота $k_{rj}^{\text{ит}}$ заданного числа K, r_j вырабатывает решение в отношении альтернативы A_{ij}. Вероятность события $P(A_{ij} = A_{opt})$ зависит от количества плиток разного цвета, встре-

тившихся роботу на пути и погрешности бортовых сенсорных устройств. В качестве альтернативы в сценариях коллективного восприятия может рассматриваться цвет плитки, доминирующий цвет плиток на сцене, процентное соотношение цветовой гаммы, локальная карта сцены и тому подобное, в зависимости от задачи, стоящей перед МРТС. Коллективное решение задачи предполагает обмен альтернативами в процессе обследования сцены [6, 8, 9].

В настоящей работе рассматривается один из сценариев коллективного восприятия — картографирование. Роботы должны исследовать свою среду (сцену), состоящую из плиток 5 цветов, и коллективно принять решение, какой образ исследуемой среды наиболее близок к оригиналу. При картографировании исследуемой сцены, состоящей из 100 плиток, роботы составляют локальные карты обследованных участков, из которых затем составляется глобальная карта сцены (рис. 1).

Из рис. 1 видно, что при случайных траекториях движения агентов роя возможны ситуации, когда некоторые плитки будут посещены несколькими роботами. В случае, когда решение робота r_j в отношении альтернативы A_i совпадает с решением робота r_k в отношении этой же альтернативы ($A_{ij} = A_{ik}$), считается, что консенсус достигнут, и $A_i = A_{opt}$. В противном случае методы DMVD и DMMD предусматривают использование определенных механизмов достижения консенсуса. Так при использовании метода DMVD рой принимает альтернативу, выдвинутую случайно выбранными агентами — выборщиками. В DMMD консенсус достигается на основе «мнения» большинства. При этом используются адаптивные взвешенные веса на основе алгоритма Метрополиса–Хастинга для оптимальной диффузии структурированных сообщений через сеть агентов.

Напомним, что к особенностям МРТС относятся такие факторы как ограниченные возможности агентов в области восприятия, вычислений и коммуникации. В рамках сценариев (рис. 1) агенты могут общаться только с другими роботами в радиусе действия связи, способны исключительно определять цвет плиток непосредственно под ними и обладают простым реактивным поведением в отношении исследуемой среды.

Перечисленные факторы можно расценивать как уязвимости, способствующие реализации угроз нарушения конфиденциальности и целостности информации, циркулирующей в МРТС. Источниками угроз для МРТС могут быть вредоносные роботы (ВР), внедренные в рой и осуществляющие деструктивное воздействие на информацию [10, 11].

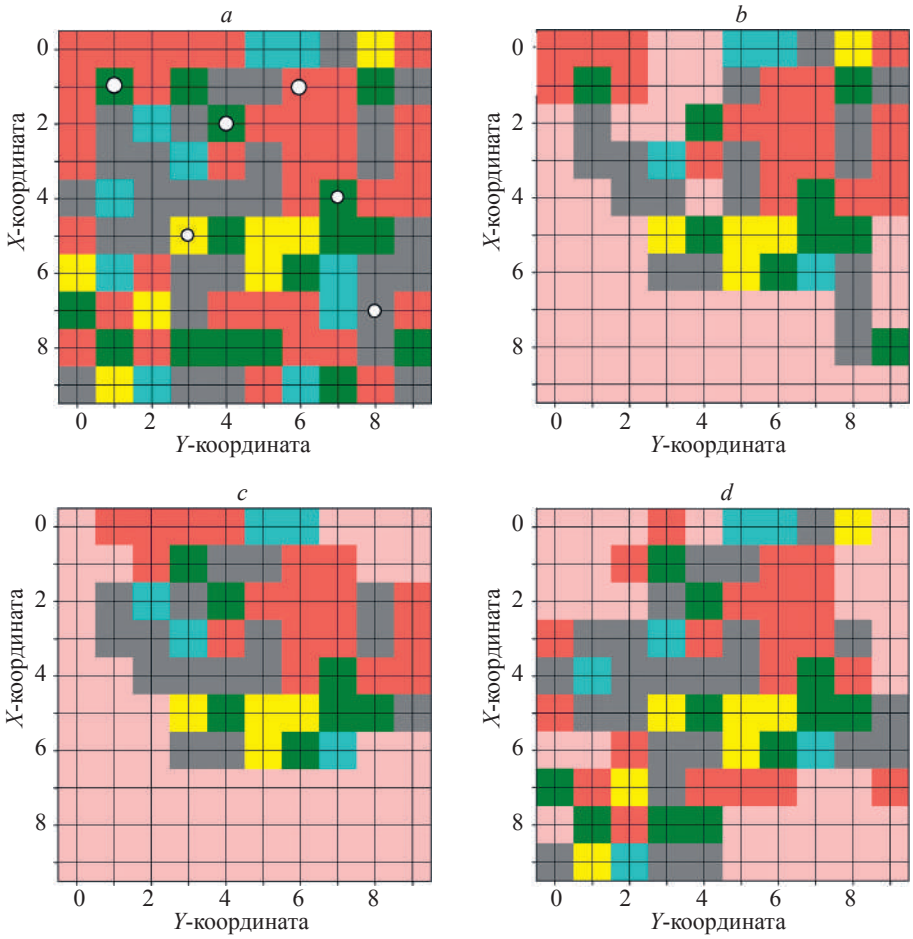

Рис. 1. Исходная сцена с 6 роботами (белые круги) (*a*) и локальные карты сцен, составленные тремя исправными роботами 1–3 роя (*b–d*)

Fig. 1. Original scene with 6 robots (white circles) (*a*) and local scene maps built by three functional robots 1–3 of the swarm (*b–d*)

В известных научных работах описаны три основные атаки ВР [10]: со случайной стратегией поведения (ССП), с оппозиционной стратегией поведения (ОСП), с координированной стратегией поведения (КСП). Суть ССП состоит в том, что каждый ВР на каждой итерации процедуры исследования среды предлагает альтернативу A^* из множества L доступных, выбранную случайным образом. ВР с ОСП предлагает любую A^{**}, но которая никогда не совпадает с истинной — A_{opt}. Главное отличие КСП от других стратегий ВР заключается в том, что все ВР с КСП изначально имеют глобальную предустановку, и выбор на каждой итерации процесса достижения консенсуса происходит в пользу некой альтернативы A^* всеми ВР с КСП [10]. Здесь $A^* \in \{A_1, A_2, \ldots, A_L\}$, $A_{opt} \in \{A_1, A_2, \ldots, A_L\}$, $A^{**} \neq A_{opt}$.

На рис. 2 показаны результаты атак ВР на рой при составлении ими локальной карты исходной сцены (рис. 1).

При сравнении исходной сцены (рис. 1) с локальной картой ВР видно, что ни одна локальная карта не идентична соответствующему фрагменту исходной сцены. При использовании ВР ОСП (робот 5) цвет ни одной из плиток локальной карты не соответствует реальному. При стратегии ССП (робот 6) возможны только случайные совпадения, а при КСП (робот 4) все плитки локальной карты представлены одним цветом. Очевидно, что при составлении глобальной карты сцены на основе локальной карты ВР возможны следующие ситуации.

1) Глобальная карта будет составлена заведомо ошибочно, если на данном участке местности был только ВР с недостоверной локальной картой.
2) Один и тот же участок сцены посетили исправные роботы (ИР) и ВР. При сопоставлении представленных ими локальных карт возникает конфликт с неопределенным исходом в отношении определения достоверности информации об исходной сцене. При этом, если для данного участка ВР обеспечат локальное большинство голосов, то консенсус будет принят на основе дезинформации ВР, и $A_i \neq A_{opt}$.

Отметим, что описанные механизмы достижения консенсуса, используемые в DMVD, DMMD и других агентно-ориентированных методах, не способны выявить подобные атаки, так как выявление ВР протоколами агентно-ориентированных методов не предусмотрено.

Таким образом, задача состоит в разработке такого метода, который даст возможность выявлять ВР даже тем ИР, которые картографируют другие участки сцены и собственными сенсорами не могут подтвердить или опровергнуть поступающую от агентов через радиосеть роя информацию или дезинформацию о свойствах исходной сцены. В этом случае ВР не смогут обеспечить локальное большинство голосов при выработке консенсуса для любого локального участка сцены.

Предлагаемое решение

Существует два принципиально разных подхода к обеспечению информационной безопасности в МРТС. Первый подход предусматривает наличие в МРТС элементов централизованного управления безопасностью роя [12–14]. Эти методы не являлись предметом исследования настоящей работы.

При втором подходе функции обеспечения информационной безопасности в гомогенных МРТС возлагаются непосредственно на агентов роя. Он включает методы защиты, основанные на введении метрик доверия и репутации агентов [15, 16], на основе технологий распределенного реестра [17], путем вычисления степени уверенности [10] и другие подходы, например основанные на использовании байесовского решающего правила [8, 18]. Однако все эти механизмы оказываются не эффективными в гомогенных системах, если при голосовании в конфликтной ситуации в отношении какой-либо альтернативы число голосов ВР превышает число голосов ИР, посетивших этот участок.

В основу предлагаемого метода положена гипотеза, что все участки исследуемой сцены обладают схожими количественными характеристиками распределения цветов. Тогда каждый робот анализирует сообщения, поступающие от всех остальных роботов, и на основе принятого в рое решающего правила оценивает инфор-

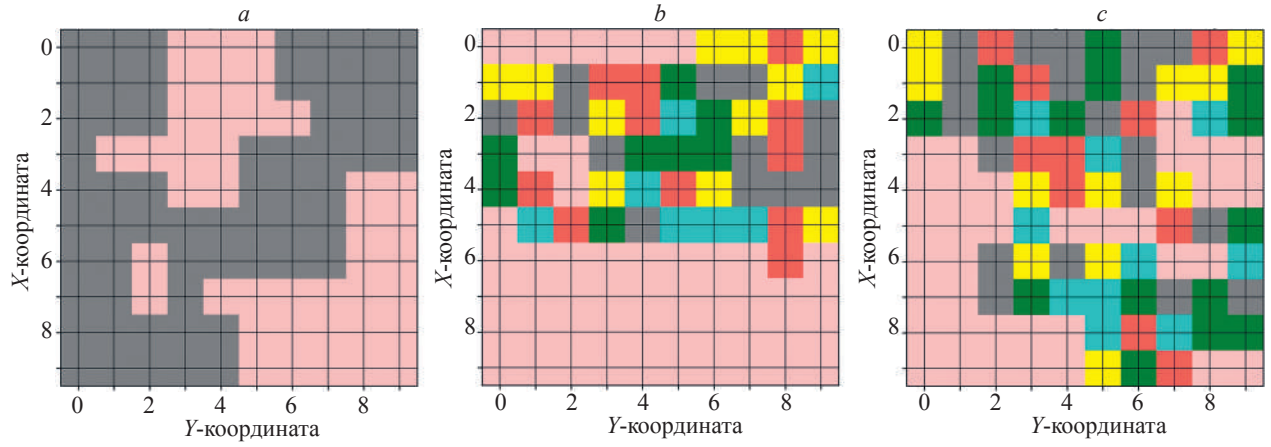

Рис. 2. Локальные карты вредоносных роботов с координированной (робот 4) (*a*), оппозиционной (робот 5) (*b*) и случайной (робот 6) (*c*) стратегиями поведения

Fig. 2. Local maps of malicious robots with coordinated (robot 4) (*a*), oppositional (robot 5) (*b*), and random (robot 6) (*c*) behavior strategies

мацию, содержащуюся в локальной карте о свойствах исследованных участков сцены и, сравнивает ее со свойствами, характерными для всей сцены. На основании этой оценки принимает решение, заслуживает ли доверия информация, переданная текущим агентом, или нет. Все локальные карты роботов, которых классификатор оценил как не относящиеся к классу ИР, из составления глобального образа сцены исключаются.

Рассмотрено три типа сцен.

Тип 1 («окрашенный шум») — с малым радиусом корреляции двумерной автокорреляционной функции (ДАКФ) (относительно размеров сцены) при наличии доминирующих цветов: 40 % плиток окрашено в цвет 1, 30 % — в цвет 2, по 10 % плиток окрашено в цвета 3–5.

Тип 2 («белый шум») — сцена с малым радиусом корреляции и равномерным распределением цветов по плиткам сцены.

Тип 3 — сцена с большим радиусом корреляции ДАКФ, сопоставимым с размерами сцены.

В качестве характеристики исходной сцены приняты гистограммы — вектор чисел q_i, i =1, 2, ..., 5, описывающих процентное соотношение плиток различных цветов: $\mathbf{hist}_{opt} = (q_1, q_2, q_3, q_4, q_5)$. Тогда гистограмма локальной карты, составленной агентом r_j, будет обозначена: $\mathbf{hist}_j = (\hat{q}_{j1}, \hat{q}_{j2}, \hat{q}_{j3}, \hat{q}_{j4}, \hat{q}_{j5})$, где $\hat{q}_{ij}$ — оценка j-м роботом величины q_i. Более сложные характеристики в настоящей работе не рассматривались ввиду ограниченной характеристики бортовых вычислительных ресурсов агентов роя. Тогда, согласно сформулированной гипотезе, должно быть верным выражение:

$$\mathbf{hist}_{opt} \approx \mathbf{hist}_{\text{ИР}}, \quad \mathbf{hist}_{opt} \neq \mathbf{hist}_{\text{ВР}},$$

где $\mathbf{hist}_{\text{ИР}}$ и $\mathbf{hist}_{\text{ВР}}$ — гистограммы локальных карт ИР и ВР.

На разработанном программном комплексе на языке Python 3.9 проведен эксперимент, в рамках которого были сформированы сцены для трех типов. Для формирования сцен типов 1 и 2 использовалась генерация случайных чисел с различным распределением плиток по цветам — с заданным (тип 1) и равномерным (тип 2) процентными соотношениями. Для формирования сцен типа 3 применен генератор чисел с нормальным распределением и последующей гауссовой фильтрацией для создания пространственной корреляции. Величина пространственной корреляции задавалась подбором значения среднего квадратического отклонения.

На исходной сцене случайным образом были расположены 6 роботов (Р1, Р2, Р3, Р4, Р5 и Р6), три из которых являлись ВР с разными стратегиями поведения (ОСП, ССП и КСП). Запускалась итерационная процедура исследования сцены, в рамках которой роботы поочередно перемещались на соседнюю свободную плитку, и определяли бортовыми сенсорами ее цвет, формируя таким образом свою локальную карту (рис. 1 и 2). Если соседние плитки оказывались заняты другими роботами, то ход пропускался. Путем подсчета количества плиток различных цветов, встреченных роботами, на каждой итерации составлялись гистограммы цветов для каждого робота в процентах от общего количества исследованных плиток. Распределение цветов в гистограммах, сформированных по окончании итерационной процедуры, представлены на рис. 3 в виде столбчатых диаграмм.

На рис. 3 видно, что гистограммы $\mathbf{hist}_{\text{ИР}}$ и $\mathbf{hist}_{\text{ВР}}$ роботов с различными стратегиями поведения могут иметь как существенное сходство, так и различие в зависимости от типа сцены. Для определения связи между гистограммами ИР и ВР на каждой итерации вычислялся коэффициент корреляции Пирсона для некоторых пар агентов

На рис. 4 показаны результаты вычислений, полученные при проведении эксперимента на сцене рис. 1 при составлении локальных карт, показанных на рис. 1 и 2. Финальная гистограмма цветов в этом эксперименте показана на рис. 3, *с*.

Как видно из графиков (рис. 4) коэффициент корреляции гистограмм ИР уже после 50 итераций достигает значение более 0,9, в то время как для пар с участием ВР соответствующие значения лежат в пределах от –0,4 до 0,74.

Для принятия решения том, является ли i-й агент ВР или ИР, использован наивный байесовский классификатор, в котором класс объекта $\hat{c}$ выбирается по правилу максимума апостериорной вероятности:

$$\hat{c} = \operatorname{argmax}[\ln P(c) + \sum_{i=1}^{n} \ln P(\mathbf{hist}_i/c)],$$

где $P(c)$ — априорная вероятность класса; $P(\mathbf{hist}_i/c)$ — плотность распределения вектора признаков i-й агента $\mathbf{hist}_i$ при условии, что он принадлежит классу c.

В качестве допущений примем следующее:

1) все признаки (количество плиток разных цветов в локальной карте робота) являются независимыми случайными величинами;
2) $P(x_i/c) \cong N(m_c, \sigma_c)$, где m_c и σ_c — математическое ожидание и среднеквадратическое отклонение случайной величины.

Выбор байесовского решающего правила обусловлен относительной простотой его реализации по сравнению с искусственными нейронными сетями, методом опорных векторов и другими классификаторами, что является желаемым свойством при реализации в бортовых вычислительных устройствах роботов.

Для оценки качества предложенного решения проведен эксперимент на сцене рис. 1, суть которого заключалась в генерации 200 агентов роя, из которых 100 относились к классу 1 (ИР), и 100 — к классу 2 (ВР). В составе выборки ВР имелось: 33 агента с ОСП, 33 — с КСП и 34 — с ССП. В рамках эксперимента была осуществлена разметка данных и выполнено распознавание агентов с использованием процедуры кросс-валидации. Результаты классификации представлены на рис. 5.

Как видно из результатов эксперимента, разработанный метод позволяет исключить из процедуры коллективного восприятия сцены с доминирующим цветом с вероятностью, приближающейся к 1. Необходимо отметить, что классификация осуществлялась на основе статистики цветов локальной карты всех роботов, независимо от того, какую часть сцены они исследовали.

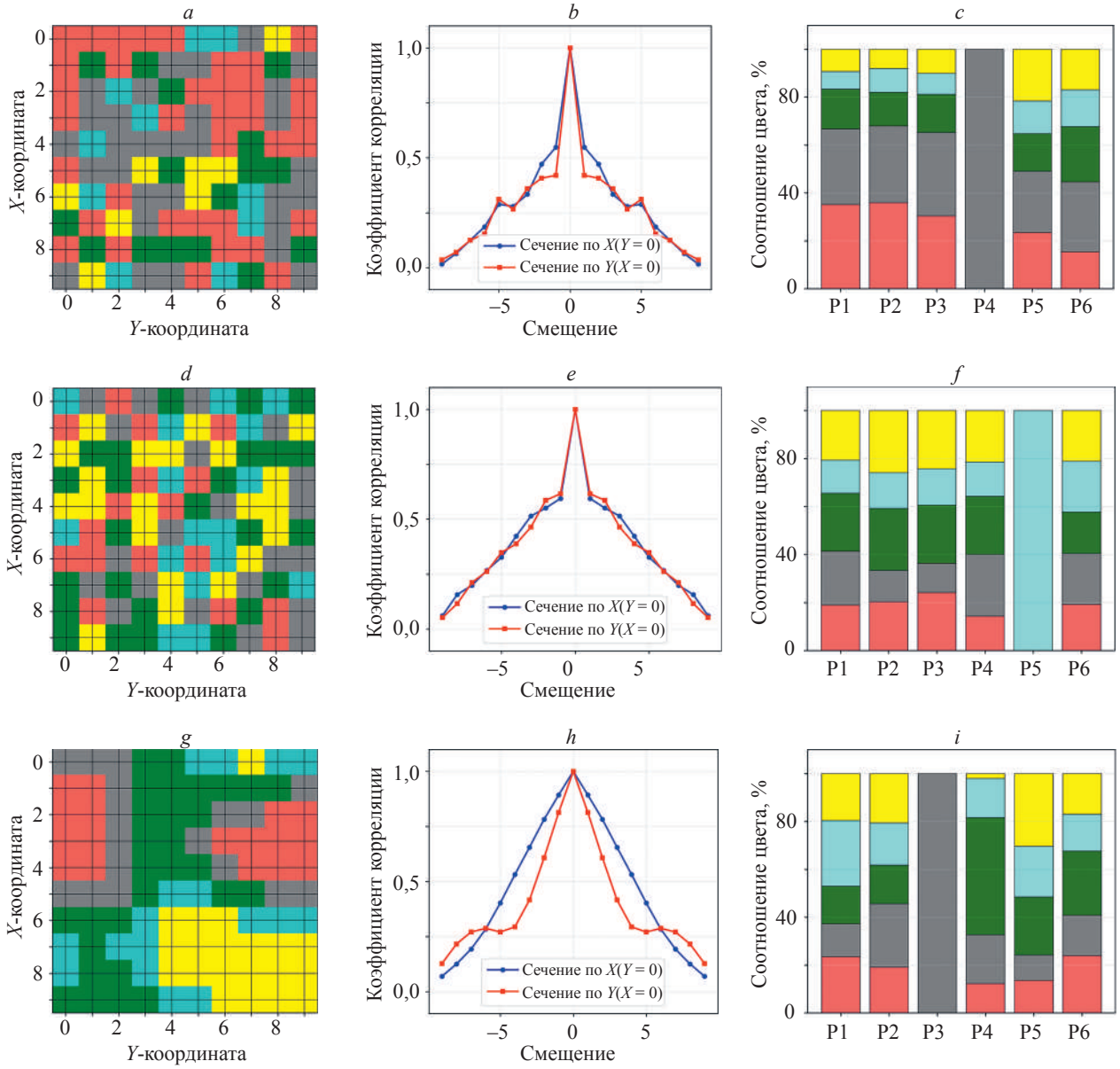

Рис. 3. Исходные сцены (*a*, *d*, *g*); сравнение сечений нормированных двумерных автокорреляционных функций по осям *X* и *Y* (*b*, *e*, *h*) и столбчатые диаграммы цветов для исправных роботов (P1–P3) и вредоносных роботов для координированной (P4), оппозиционной (P5) и случайной (P6) стратегий поведения (*c*, *f*, *i*). Радиусы корреляций сцен: для типов 1 (*a*–*c*) и 2 (*d*–*f*) (около 1), для типа 3 (*g*–*i*) (примерно 5)

Fig. 3. Original scenes (*a*, *d*, *g*); comparison of normalized 2D autocorrelation function cross-sections along the *X* and *Y* axes (*b*, *e*, *h*); and color bar charts for functional robots (P1–P3) and malicious robots with coordinated (P4), oppositional (P5), and random (P6) behavior strategies (*c*, *f*, *i*). Scene correlation radii: for types 1 (*a*–*c*) and 2 (*d*–*f*) (about 1), for type 3 (*g*–*i*) (approximately 5)

Это требование было положено в основу предложенного метода.

Картина существенно изменилась при проведении такого эксперимента на сцене с большим радиусом корреляции (рис. 3, *g*–*i*), результаты которого показаны на рис. 6.

Как видно из представленных результатов, имеет место большая ошибка классификации второго рода, обусловленная низкой точностью распознавания объектов класса 2, которая является неприемлемой для решения задачи выявления ВР. Очевидно, что атрибуты классификации, основанные на подсчете частоты различных цветов плиток локальной карты роботов, не позволяют учесть величину пространственной корреляции объектов сцены. Иначе говоря, несмотря на очевидную разницу в представлении локальной карты сцены с большим радиусом корреляции (рис. 6), предложенное решение не обеспечило требуемое качество.

Как видно из рис. 7, изображение локальной карты ВР представляет собой более фрагментированную картину, в отличие от ИР. В связи с этим был введен дополнительный атрибут классификации, позволяющий оценивать пространственную протяженность объек-

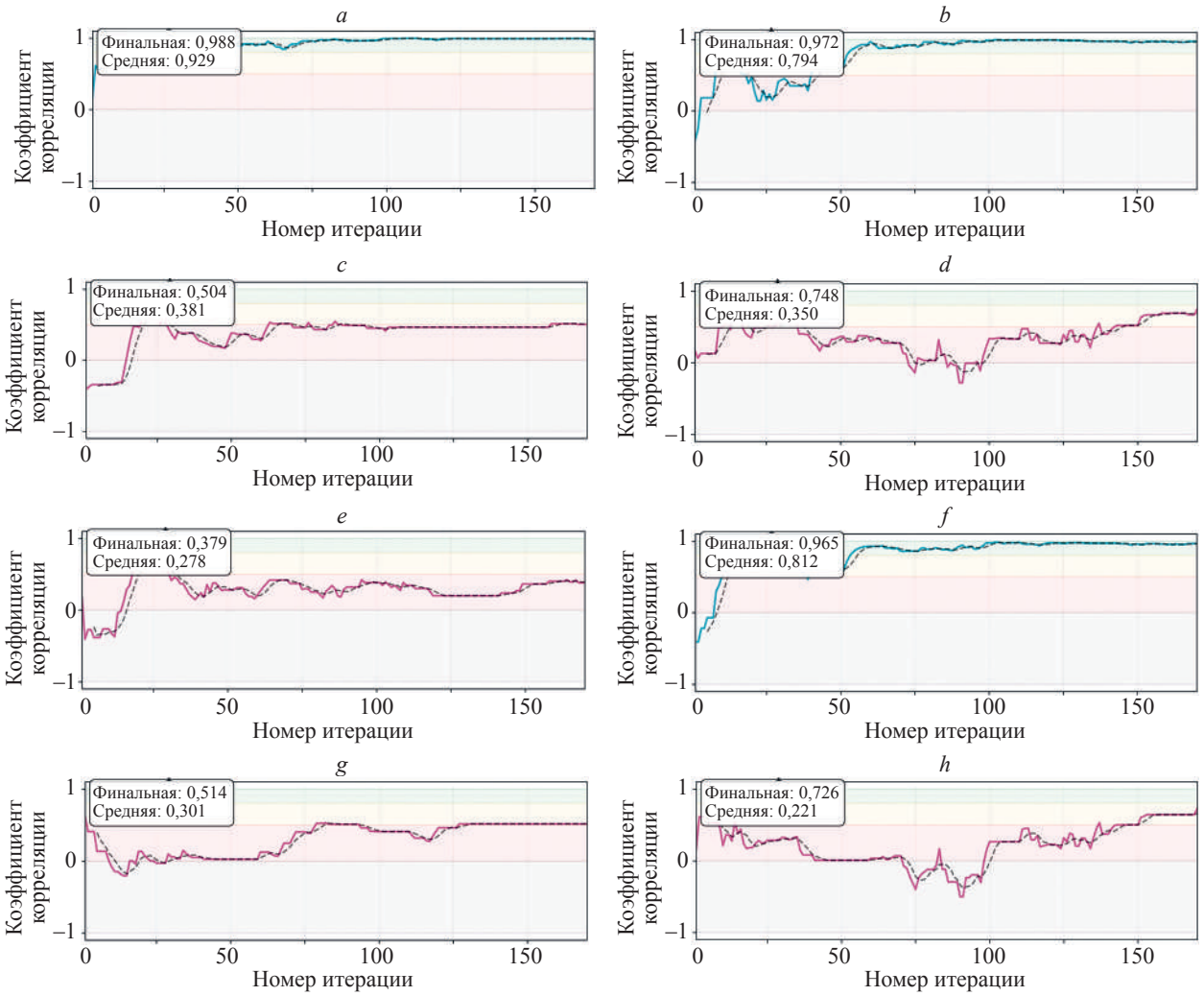

Рис. 4. Зависимости коэффициентов корреляции между парами роботов: 1 и 2 (*a*), 1 и 3 (*b*), 1 и 4 (*c*), 1 и 5 (*d*), 1 и 6 (*e*), 2 и 3 (*f*), 2 и 4 (*g*), 2 и 5 (*h*), полученные экспериментально для сцены типа 1 для 5 признаков.

Гистограммы цветов: финальная (сплошная линия) и средняя (пунктирная линия)

Fig. 4. Dependences of correlation coefficients between pairs of robots: 1 and 2 (*a*), 1 and 3 (*b*), 1 and 4 (*c*), 1 and 5 (*d*), 1 and 6 (*e*), 2 and 3 (*f*), 2 and 4 (*g*), 2 and 5 (*h*) obtained experimentally for a type 1 scene for 5 signs.

Color histograms: final (solid line) and average (dashed line)

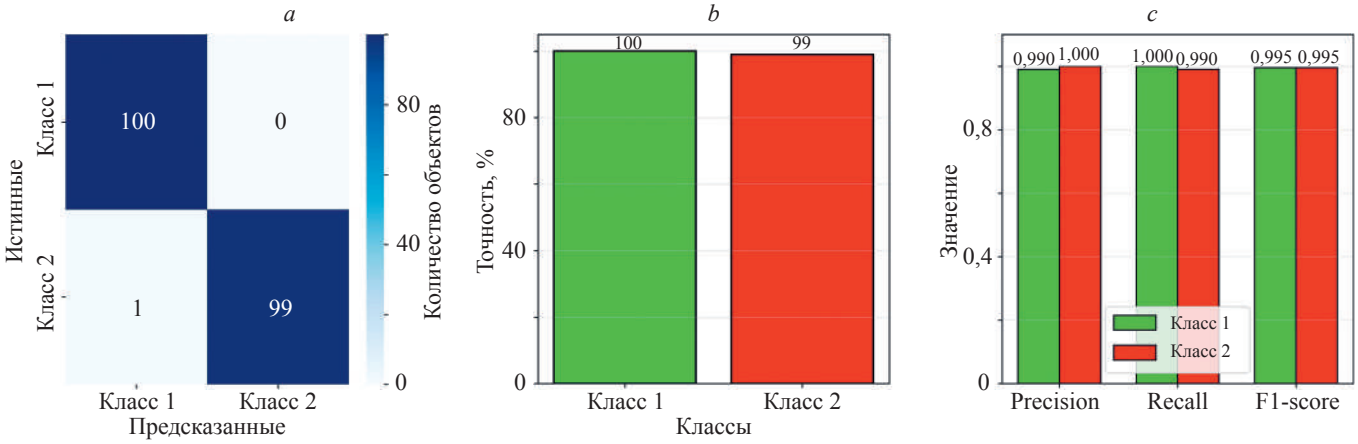

Рис. 5. Показатели качества классификации 200 агентов, исследующих сцену типа 1: матрица ошибок классификации (*a*); точность классификации по классам (*b*); сравнение метрик по классам (*c*)

Fig. 5. Classification quality metrics for 200 agents exploring a Type 1 scene: classification error matrix (*a*); per-class classification accuracy (*b*); class-wise metric comparison (*c*)

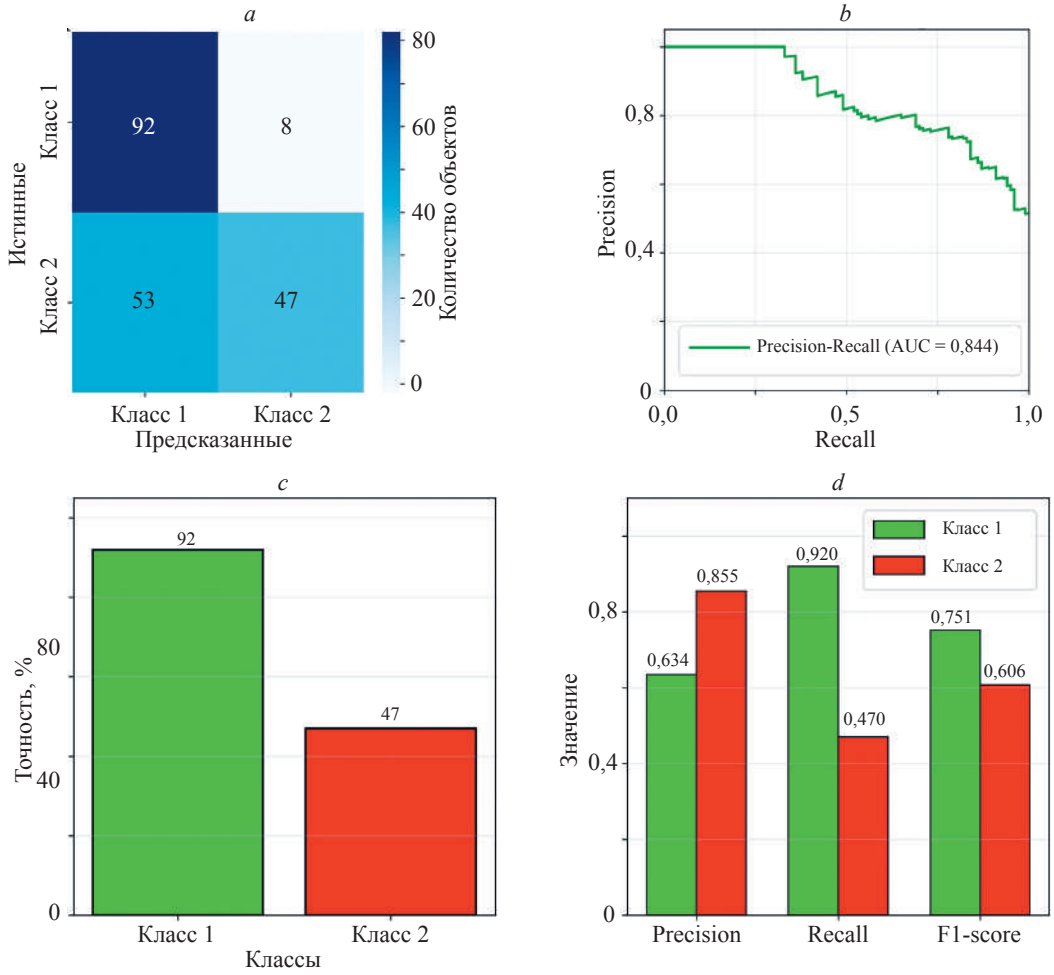

Рис. 6. Показатели качества классификации 200 агентов, исследующих сцену типа 3 с большим радиусом корреляции: матрица ошибок классификации (*a*); Precision-Recall-кривая (*b*); точность классификации (*c*) и сравнение метрик (*d*) по классам

Fig. 6. Classification quality metrics for 200 agents exploring a Type 3 scene with a large correlation radius: classification error matrix (*a*); Precision-Recall curve (*b*); classification accuracy (*c*); and class-wise metric comparison (*d*)

тов на сцене. Чтобы не прибегать к дополнительным мощностям бортовых устройств в качестве дополнительного атрибута был принят процент одноцветных смежных плиток в составе локальной карты. Подсчет этой величины существенно проще, чем вычисление ДАКФ.

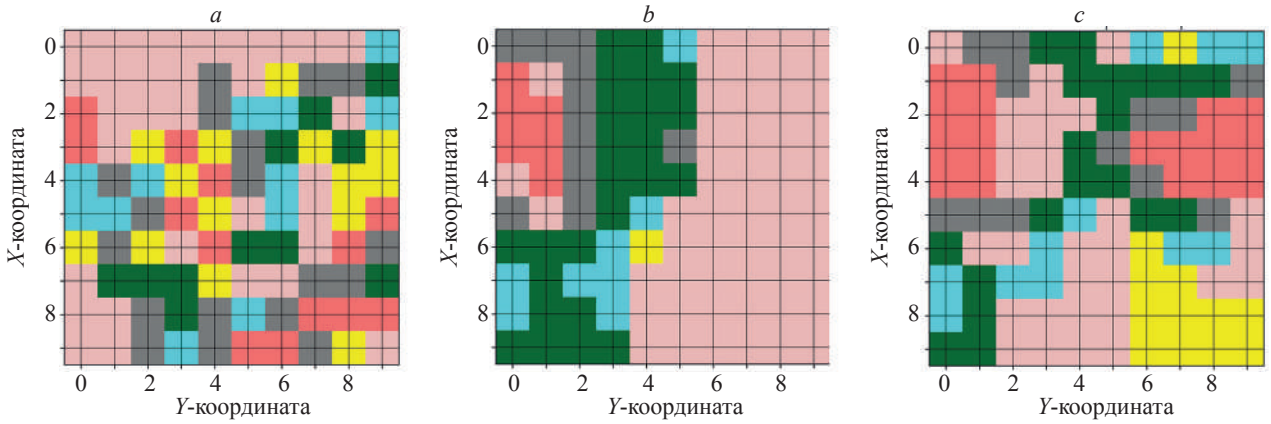

Рис. 7. Локальные карты роботов: вредоносного (робот 2) (*a*); исправных (робот 4) (*b*) и (робот 6) (*c*) при исследовании сцены с радиусом корреляции около 5 плиток

Fig. 7. Local robot maps: malicious (robot 2) (*a*); operational (robot 4) (*b*) and (robot 6) (*c*) when exploring a scene with a correlation radius of about 5 tiles

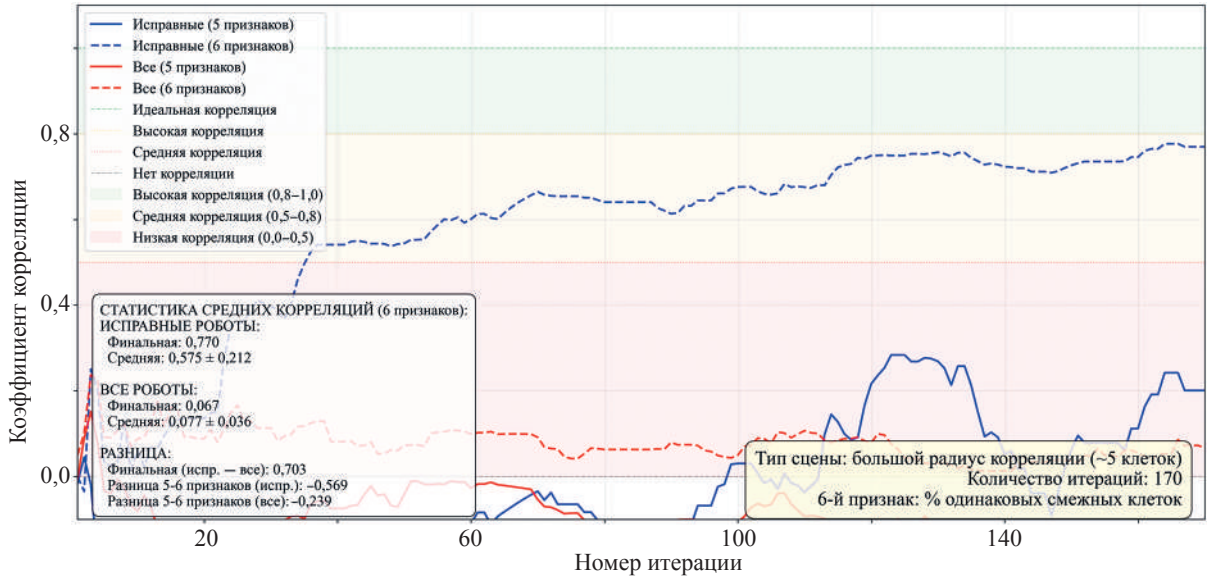

Рис. 8. Зависимость средней корреляции от количества итераций при 5 и 6 признаках классификации. Тип сцены: большой радиус корреляций

Fig. 8. Dependence of the average correlation on the number of iterations for 5 and 6 classification features. Scene type: large correlation radius

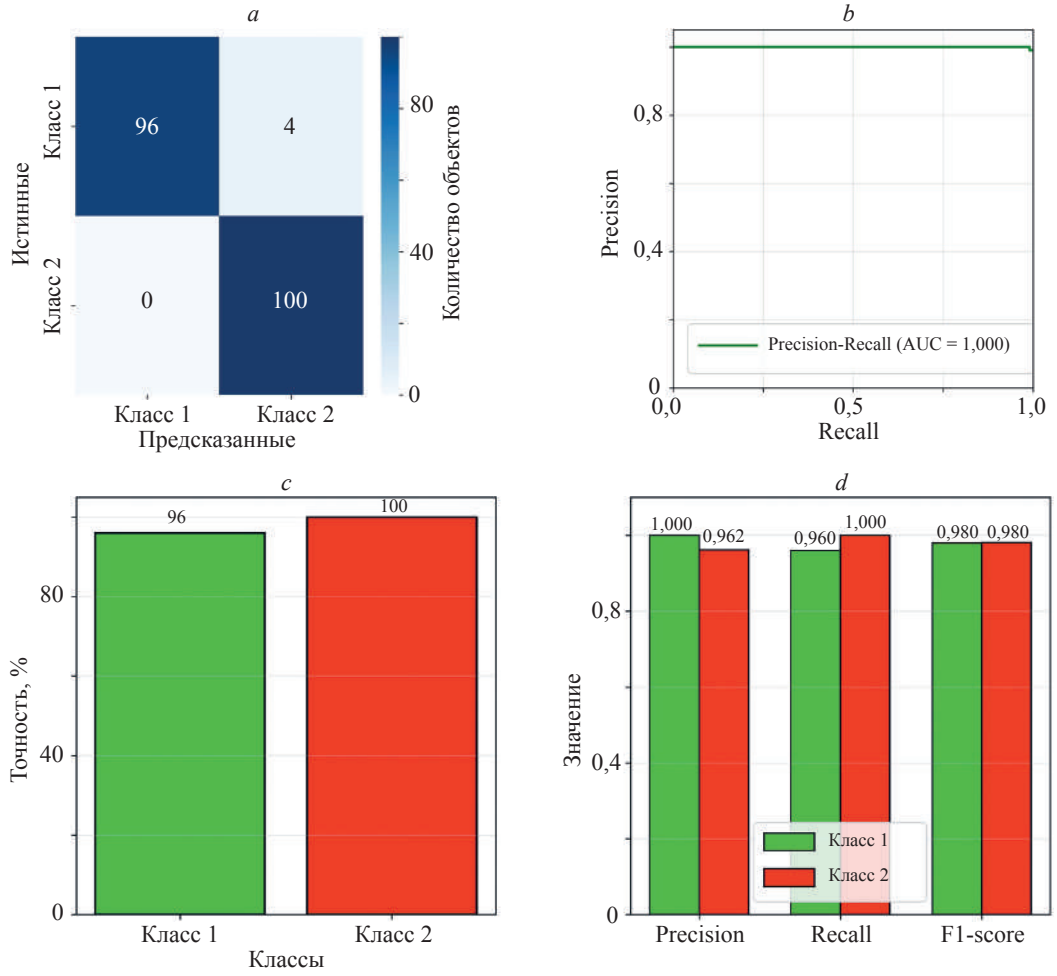

Рис. 9. Показатели качества классификации 200 агентов, исследующих сцену типа 3 с большим радиусом корреляции с использованием дополнительного признака распознаваемого объекта: матрица ошибок классификации (*a*); Precision-Recall-кривая (*b*); точность классификации (*c*) и сравнение метрик (*d*) по классам

Fig. 9. Classification quality metrics for 200 agents exploring a Type 3 scene with a large correlation radius using an additional feature of the recognized object: classification error matrix (*a*); Precision-Recall curve (*b*); classification accuracy (*c*); and class-wise metric comparison (*d*)

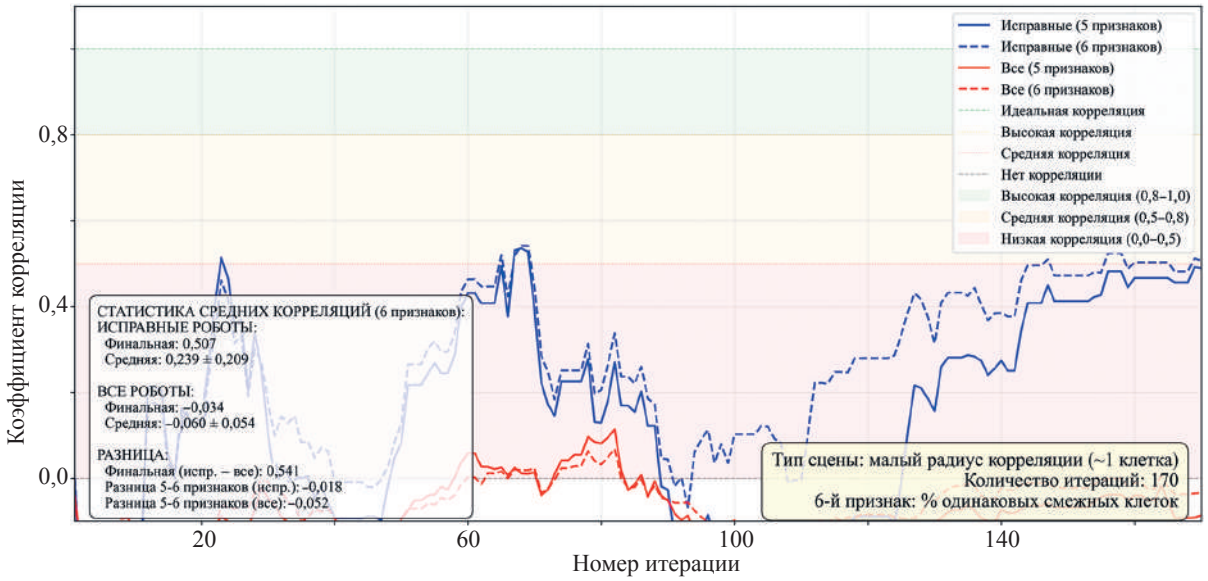

Рис. 10. Зависимость средней корреляции от количества итераций. Тип сцены: малый радиус корреляций

Fig. 10. Dependence of the average correlation on the number of iterations. Scene type: small correlation radius

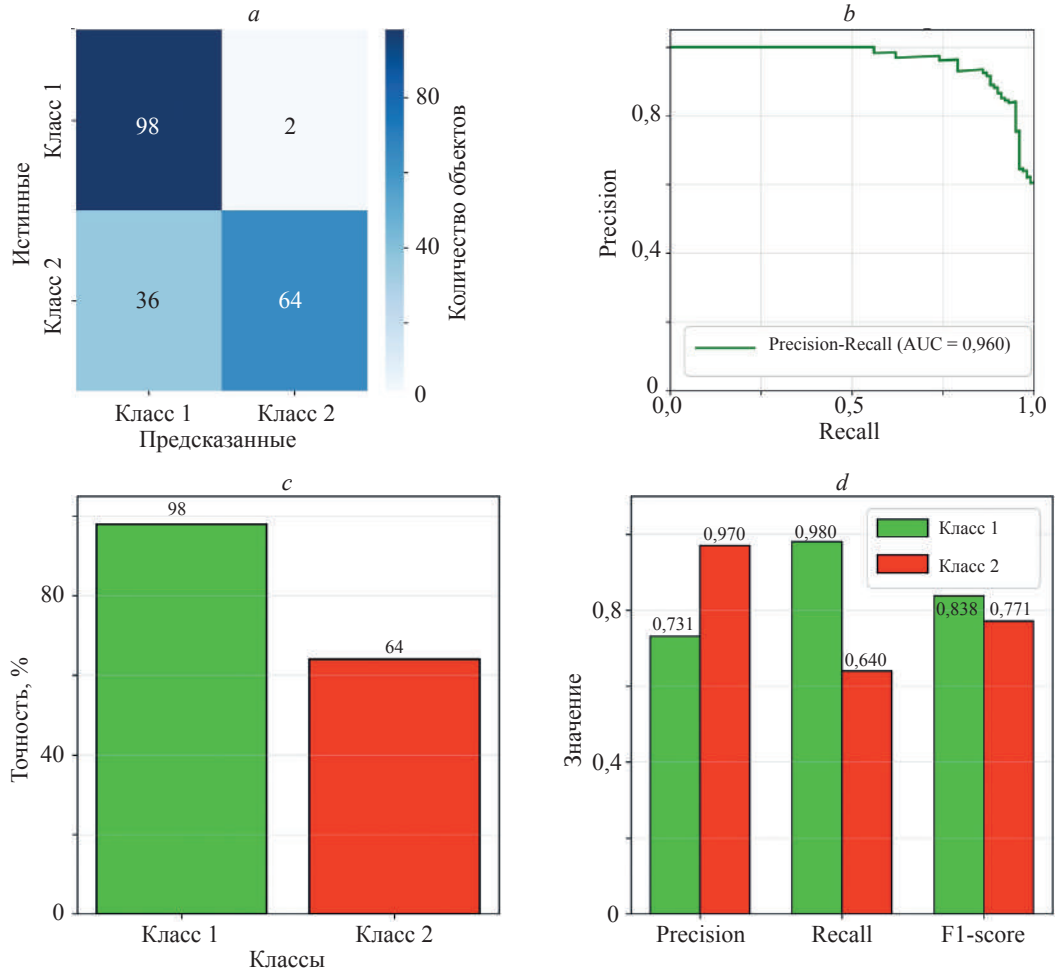

Рис. 11. Показатели качества классификации 200 агентов, исследующих сцену типа 2: матрица ошибок классификации (*a*); Precision-Recall-кривая (*b*); точность классификации (*c*) и сравнение метрик (*d*) по классам

Fig. 11. Classification quality metrics for 200 agents exploring a Type 2 scene: classification error matrix (*a*); Precision-Recall curve (*b*); classification accuracy (*c*); and class-wise metric comparison (*d*)

Введение в состав гистограмм $\mathbf{hist}_{opt}$, $\mathbf{hist}_{ИР}$ и $\mathbf{hist}_{ВР}$ дополнительного признака g_6, описывающего пространственную протяженность объектов на сцене, привело к существенному увеличению степени взаимосвязи между $\mathbf{hist}_{ИР}$ по сравнению с корреляцией гистограмм 5 признаков (рис. 8).

Из рис. 8 видно, что с увеличением числа итераций коэффициент корреляции между ИР приближается к 0,9, что дает основания полагать о возможности распознавания ВР.

Результаты классификации для тех же исходных данных с использованием 6 атрибутов классификатора представлены на рис. 9.

Как следует из проведенного эксперимента, использование в качестве дополнительного атрибута классификатора простейшего анализатора пространственной протяженности объектов на сцене существенно улучшило качество распознавания. Удалось минимизировать ошибку второго рода и обеспечить 100 % точность выявления объектов класса 2 — ВР.

Иная ситуация возникает при исследовании сцены типа 2. При равномерном распределении цветов с малым радиусом корреляции ДАКФ (примерно 1 плитка) отсутствует какая-либо «сигнальная» составляющая в частотной характеристике сцены. Результаты расчета подтвердили отсутствие корреляционной связи между локальной картой ИР как при 5, так и при 6 атрибутах в гистограмме цветов агентов (рис. 10).

Как видно из результатов расчетов средней корреляции (рис. 10) между ИР и ВР, увеличение количества итераций не обеспечивает достижение высокой степени корреляции между гистограммами роботов при исследовании сцены типа 2.

Результаты эксперимента для 200 объектов по картографированию сцены типа 2 представлены на рис. 11.

Таким образом, предлагаемое решение обеспечивает «штатную» работу алгоритмов DMMD, DMVD, и Direct Comparison без влияния ВР при картографировании сцен с цветовыми или пространственными паттернами, что соответствует постановке задачи настоящей работы.

Заключение

В работе расширены возможности коллективного принятия решений до нового сценария дискретной коллективной оценки с различными вариантами исходных данных. Использованы три типа сцен в задаче коллективного восприятия исходной сцены с целью ее картографирования. Исследования проводились на сцене с малым радиусом двумерной автокорреляционной функции при наличии доминирующих признаков и равномерным распределением признаков, и на сцене с большим радиусом двумерной автокорреляционной функции, сопоставимым с размерами сцены. В результате эксперимента применены три типа атак на рой роботов, осуществляющих картографирование исходной сцены — атаки роботов с оппозиционной случайной и консолидированной стратегиями поведения.

Сформулировано требование, что для исключения локального большинства вредоносных роботов при выработке роем консолидированного решения, метод должен позволять распознавать их всем роботам роя вне зависимости от того, имелся ли конфликт интересов между агентами. Предложены атрибуты, характеризующие деятельность каждого робота в виде частотной гистограммы цветов и количества смежных участков на картографируемой сцене. Разработанный метод основан на процедуре распознавания образов, атрибутами которой являлись частотные гистограммы локальной карты каждого робота.

Представленный метод был протестирован на трех типах сцен, отличающимися частотными и пространственными характеристиками при наличии исправных и вредоносных роботов с тремя стратегиями поведения. Показано, что для достижение высоких показателей качества распознавания обучение классификатора необходимо проводить на сцене такого же типа как картографируемый объект.

Экспериментально показано, что предлагаемый метод позволяет выявлять вредоносных роботов независимо от используемой ими стратегии поведения на таких типах сцен, которые обладают частотными или линейными паттернами.

В дальнейших исследованиях планируется упростить процедуру распознавания образов для уменьшения нагрузки на бортовые вычислительные устройства за счет снижения размерности вектора признаков классификации.

Литература

1. Sailor M.J., Link J.R. “Smart dust”: nanostructured devices in a grain of sand // Chemical Communications. 2005. V. 11. P. 1375–1385. doi: 10.1039/b417554a
2. Каляев И.А., Гайдук А.Р., Капустян С.Г. Модели и алгоритмы коллективного управления в группах роботов. М.: Физматлит, 2009. 280 с.
3. Городецкий В.И. Поведенческие модели кибер-физических систем и групповое управление: основные понятия // Известия ЮФУ. Технические науки. 2019. № 1 (203). С. 144–162. doi: 10.23683/2311-3103-2019-1-144-162
4. Valentini G., Brambilla D., Hamann H., Dorigo M. Collective perception of environmental features in a robot swarm // Lecture Notes in Computer Science. 2016. V. 9882. P. 65–76. doi: 10.1007/978-3-319-44427-7_6

References

1. Sailor M.J., Link J.R. “Smart dust”: nanostructured devices in a grain of sand. *Chemical Communications*, 2005, vol. 11, pp. 1375–1385. doi: 10.1039/b417554a
2. Kalyaev I.A., Gaiduk A.R., Kapustyan S.G. *Models and Algorithms of Collective Control in Groups of Robots*. Moscow, Fizmatlit Publ., 2009, 280 p. (in Russian)
3. Gorodetsky V.I. Behavioral model for cyber-physical system and group control: the basic concepts. *Izvestiya SFEDU. Engineering Sciences*, 2019, no. 1 (203), pp. 144–162. (in Russian). doi: 10.23683/2311-3103-2019-1-144-162
4. Valentini G., Brambilla D., Hamann H., Dorigo M. Collective perception of environmental features in a robot swarm. *Lecture Notes in Computer Science*, 2016, vol. 9882, pp. 65–76. doi: 10.1007/978-3-319-44427-7_6

5. Valentini G., Hamann H., Dorigo M. Self-organized collective decision making: The weighted voter model // Proc. of the 13th International Conference on Autonomous Agents and Multiagent Systems. 2014. P. 45–52. doi: 10.65109/mdde7348
6. Valentini G., Ferrante E., Hamann H., Dorigo M. Collective decision with 100 Kilobots: Speed versus accuracy in binary discrimination problems // Autonomous Agents and Multi-Agent Systems. 2015. V. 30. N 3. P. 553–580. doi: 10.1007/s10458-015-9323-3
7. Castellano C., Fortunato S., Loreto V. Statistical physics of social dynamics // Reviews of Modern Physics. 2009. V. 81. N 2. P. 591–646. doi: 10.1103/revmodphys.81.591
8. Зикратов И.А., Зикратова Т.В., Новиков Е.А. Реализация стратегии коллективного восприятия в самоорганизующейся роевой системе с использованием байесовского решающего правила // Труды учебных заведений связи. 2025. Т. 11. №3. С. 108–118. doi: 10.31854/1813-324X-2025-11-3-108-118
9. Basan A.S., Basan E.A., Makarevich O.B. Analysis of ways to secure group control for autonomous mobile robots // Proc. of the 10th International Conference on Security of Information and Networks. 2017. P. 134–139. doi: 10.1145/3136825.3136879
10. Рябцев С.С. Метод выявления вредоносных роботов на основе данных процесса коллективного принятия решений в роевых робототехнических системах // Системы управления, связи и безопасности. 2022. № 3. С. 105–137. doi: 10.24412/2410-9916-2022-3-105-137
11. Zikratov I.A., Lebedev I.S., Gurtov A.V., Kuzmich E.V. Securing swarm intellect robots with a police office model // Proc. of the IEEE 8th International Conference on Application of Information and Communication Technologies (AICT). 2014. P. 1–5. doi: 10.1109/icaict.2014.7035906
12. Guan X., Yang Y., You J. POM-a mobile agent security model against malicious hosts // Proc. of the 4th International Conference/Exhibition on High Performance Computing in the Asia-Pacific Region. 2000. V. 2. P. 1165–1166. doi: 10.1109/hpc.2000.843621
13. Page J., Zaslavsky A., Indrawan M. A buddy model of security for mobile agent communities operating in pervasive scenarios // Proc. of the 2nd Australasian Information Security Workshop. 2004. P. 17–25.
14. Зикратов И.А., Гуртов А.В., Зикратова Т.В., Козлова Е.В. Совершенствование police office model для обеспечения безопасности роевых робототехнических систем // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 5 (93). С. 99–109.
15. Schillo M., Funk P., Rovatsos M. Using trust for detecting deceitful agents in artificial societies // Applied Artificial Intelligence. 2000. V. 14. N 8. P. 825–848. doi: 10.1080/08839510050127579
16. Зикратов И.А., Викснин И.И., Зикратова Т.В., Шлыков А.А., Медведков Д.И. Модель безопасности мобильных робототехнических систем с коллективным управлением // Научно-технический вестник информационных технологий, механики и оптики. 2017. Т. 17. № 3. С. 439–449. doi: 10.17586/2226-1494-2017-17-3-439-449
17. Strobel V., Dorigo M. Blockchain technology for robot swarms: A shared knowledge and reputation management system for collective estimation // Proc. of the ANTS 2018 11th International Conference. 2018. P. 425.
18. Shan Q., Mostaghim S. Discrete collective estimation in swarm robotics with distributed Bayesian belief sharing // Swarm Intelligence. 2021. V. 15. N 4. P. 378–405. doi: 10.1007/s11721-021-00201-w

5. Valentini G., Hamann H., Dorigo M. Self-organized collective decision making: The weighted voter model. *Proc. of the 13th International Conference on Autonomous Agents and Multiagent Systems*, 2014, pp. 45–52. doi: 10.65109/mdde7348
6. Valentini G., Ferrante E., Hamann H., Dorigo M. Collective decision with 100 Kilobots: Speed versus accuracy in binary discrimination problems. *Autonomous Agents and Multi-Agent Systems*, 2015, vol. 30, no. 3, pp. 553–580. doi: 10.1007/s10458-015-9323-3
7. Castellano C., Fortunato S., Loreto V. Statistical physics of social dynamics. *Reviews of Modern Physics*, 2009, vol. 81, no. 2, pp. 591–646. doi: 10.1103/revmodphys.81.591
8. Zikratov I.A., Zikratova T.V., Novikov E.A. Implementation of collective perception strategy in a self-organizing swarm system using bayesian decision rule. *Proceedings of Telecommunication Universities*. 2025, vol. 11, no. 3, pp. 108–118. (in Russian). doi: 10.31854/1813-324X-2025-11-3-108-118
9. Basan A.S., Basan E.A., Makarevich O.B. Analysis of ways to secure group control for autonomous mobile robots. *Proc. of the 10th International Conference on Security of Information and Networks*, 2017, pp. 134–139. doi: 10.1145/3136825.3136879
10. Ryabtsev S.S. A method for detecting byzantine robots based on data from the collective decision-making process in swarm robotic systems. *Systems of Control, Communication and Security*, 2022, no. 3, pp. 105–137. (in Russian). doi: 10.24412/2410-9916-2022-3-105-137
11. Zikratov I.A., Lebedev I.S., Gurtov A.V., Kuzmich E.V. Securing swarm intellect robots with a police office model. *Proc. of the IEEE 8th International Conference on Application of Information and Communication Technologies (AICT)*, 2014, pp. 1–5. doi: 10.1109/icaict.2014.7035906
12. Guan X., Yang Y., You J. POM-a mobile agent security model against malicious hosts. *Proc. of the 4th International Conference/Exhibition on High Performance Computing in the Asia-Pacific Region*, 2000, vol. 2, pp. 1165–1166. doi: 10.1109/hpc.2000.843621
13. Page J., Zaslavsky A., Indrawan M. A buddy model of security for mobile agent communities operating in pervasive scenarios. *Proc. of the 2nd Australasian Information Security Workshop*, 2004, pp. 17–25.
14. Zikratov I.A., Gurtov A.V., Zikratova T.V., Kozlova E.V. Police office model improvement for security of swarm robotic systems. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2014, no. 5 (93), pp. 99–109. (in Russian)
15. Schillo M., Funk P., Rovatsos M. Using trust for detecting deceitful agents in artificial societies. *Applied Artificial Intelligence*, 2000, vol. 14, no. 8, pp. 825–848. doi: 10.1080/08839510050127579
16. Zikratov I.A., Viksnin I.I., Zikratova T.V., Shlykov A.A., Medvedkov D.I. Security model of mobile multi-agent robotic systems with collective management. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2017, vol. 17, no. 3, pp. 439–449. (in Russian). doi: 10.17586/2226-1494-2017-17-3-439-449
17. Strobel V., Dorigo M. Blockchain technology for robot swarms: A shared knowledge and reputation management system for collective estimation. *Proc. of the ANTS 2018 11th International Conference*, 2018, pp. 425.
18. Shan Q., Mostaghim S. Discrete collective estimation in swarm robotics with distributed Bayesian belief sharing. *Swarm Intelligence*, 2021, vol. 15, no. 4, pp. 378–405. doi: 10.1007/s11721-021-00201-w

Авторы

Зикратов Игорь Алексеевич — доктор технических наук, профессор, декан факультета, Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, Санкт-Петербург, 193232, Российская Федерация, sc 56321572700, https://orcid.org/0000-0001-9054-800X, igzikratov@yandex.ru

Зикратова Татьяна Викторовна — кандидат технических наук, доцент, Военно-морской политехнический институт ВУНЦ ВМФ «Военно-морская академия имени Адмирала флота Советского Союза Н.Г. Кузнецова», Санкт-Петербург, 197342, Российская Федерация, https://orcid.org/0000-0001-8365-658X, ztv64@yandex.ru

Authors

Igor A. Zikratov — D.Sc., Professor, Dean, The Bonch-Bruevich Saint Petersburg State University of Telecommunications (SPbSUT), Saint Petersburg, 193232, Russian Federation, sc 56321572700, https://orcid.org/0000-0001-9054-800X, igzikratov@yandex.ru

Tatiana V. Zikratova — PhD, Associate Professor, VUNTS of the Navy "Naval Academy", Saint Petersburg, 197342, Russian Federation, https://orcid.org/0000-0001-8365-658X, ztv64@yandex.ru

Статья поступила в редакцию 04.02.2026
Одобрена после рецензирования 14.04.2026
Принята к печати 19.05.2026

Received 04.02.2026
Approved after reviewing 14.04.2026
Accepted 19.05.2026