

ИНФОРМАЦИОННЫЕ  
ТЕХНОЛОГИИ  
И ТЕЛЕКОММУНИКАЦИИ

INFORMATION  
TECHNOLOGIES  
AND TELECOMMUNICATIONS

УДК 004.89

DOI: 10.17212/2782-2001-2022-3-21-36

**Метод автоматического распознавания  
девиантного поведения людей на основе  
интеграции технологий компьютерного зрения  
и управления знаниями для поддержки принятия  
решений операторами систем видеомониторинга\***

**И.А. РЯБЧИКОВ**

199178, РФ, г. Санкт-Петербург, 14-я линия Васильевского острова, 39, СПб ФИЦ  
РАН

[i.a.ryabchikov@gmail.com](mailto:i.a.ryabchikov@gmail.com)

Обеспечение безопасности в городской среде является важным направлением повышения качества жизни людей, и развитие современных интеллектуальных технологий открывает новые возможности для достижения этой цели. Применение современных интеллектуальных технологий позволяет раскрыть потенциал систем видеонаблюдения, делая возможным автоматическое распознавание опасных ситуаций в режиме реального времени для своевременного принятия мер по их устранению и оказанию помощи жертвам. Часто опасная ситуация обусловлена девиантным поведением людей – ограблением, дракой, вандализмом и т. д. Но в существующих работах, ориентированных на распознавание инцидентов девиантного поведения, ставится задача распознавания лишь кратковременных отличительных признаков, таких как удары, падения или оружие в руках у человека. При этом часто такие признаки могут отсутствовать, например, при ограблении прохожих без нападения и борьбы, поэтому задача распознавания сложных протяженных во времени сцен девиантного поведения остается нерешенной. В настоящей работе предложен метод автоматического распознавания протяженных сцен девиантного поведения людей, отличающийся интеграцией технологий управления знаниями и компьютерного зрения для детектирования и сегментации объектов, оценки трехмерного скелета человека, отслеживания объектов на кадрах видеозаписи, оценки нормали к поверхности земли для определения расстояния от объектов до камеры и классификации элементарных действий людей по трехмерному скелету. Данный метод может быть использован при разработке системы поддержки принятия решений операторами систем видеомониторинга, применяемой для детектирования и обработки девиантного поведения людей в режиме реального времени с целью предотвращения эскалации, оказания своевременной помощи жертвам и задержания подозреваемых.

---

\* Статья получена 12 июля 2022 г.

**Ключевые слова:** база знаний, бесхозный предмет, девиантное поведение, классификация действий, компьютерное зрение, ограбление, поддержка принятия решений, распознавание сцен, система видеонаблюдения, трехмерный скелет

## ВВЕДЕНИЕ

Одной из основных задач умного города является обеспечение безопасности жителей, и огромный потенциал для ее решения имеют системы видеонаблюдения. Они способны обеспечивать осведомленность о происходящих в городе опасных ситуациях, таких как аварии, пожары, несчастные случаи и преступления. Часто опасные ситуации обусловлены девиантным поведением людей – действиями, которые нарушают общественные нормы или права людей и которые могут привести к опасности для жизни и материальному ущербу (ограбление, драка, вандализм и т. д.). И оперативное детектирование таких ситуаций может позволить предотвратить дальнейшую эскалацию, оказать своевременную помощь жертвам и задержать подозреваемых. Но основное применение систем видеонаблюдения на данный момент – расследование происшествий уже после их завершения. Причиной этому является огромное количество камер видеонаблюдения, из-за чего оперативная ручная обработка материалов операторами систем видеонаблюдения не представляется возможной. На текущий момент в России размещено более 15 млн камер видеонаблюдения<sup>1</sup>, при этом в Москве размещено более 200 тыс. камер<sup>2</sup>, а в Санкт-Петербурге – более 58 тыс.<sup>3</sup>

Для оперативного обнаружения и противодействия опасным ситуациям применяются системы поддержки принятия решений (СППР), осуществляющие фильтрацию видеопотоков со множества камер видеонаблюдения и предоставляющие оператору в режиме реального времени информацию о детектированных инцидентах. На основе этой информации оператор (лицо, принимающее решения – ЛПР) может быстро провалидировать детектированный инцидент и принять меры по его обработке, в частности, скоординировать службы реагирования. Примерами таких систем являются системы детектирования пожаров и задымлений [1], дорожных инцидентов [2], оставленных предметов в публичных местах (Reco3.26<sup>4</sup>) и разыскиваемых преступников (FindFace<sup>5</sup>).

На данный момент существуют работы, направленные на поддержку принятия решений при детектировании девиантного поведения, но в основном они ограничены распознаванием лишь кратковременных сцен, содержащих явные признаки: удары, падения или оружие в руках у человека. В таких системах

---

<sup>1</sup> TelecomDaily: Россия вышла на второе место по темпам подключения видеокамер. Доступен по: <https://telecomdaily.ru/news/2021/06/30/telecomdaily-rossiya-vyshla-na-vtoroe-mesto-po-temпам-podklyucheniya-videokamer> (Дата обращения: 17.07.2022).

<sup>2</sup> Городская система видеонаблюдения. Доступен по: <https://video.dit.mos.ru/> (дата обращения: 17.07.2022).

<sup>3</sup> Городской мониторинговый центр. Доступен по: <http://spb112.ru/catalogue/4/> (дата обращения: 17.07.2022).

<sup>4</sup> Smart Recognition System Reco3.26. Доступен по: <https://www.reco326.com/en/> (дата обращения: 17.07.2022).

<sup>5</sup> FindFace Public Safety. Доступен по: <https://ntechlab.ru/solution/public-safety/> (дата обращения: 17.07.2022).

обычно осуществляется извлечение базовых признаков из кадров (признаков формы сегментов объектов, признаков движения, скорости и ускорения сегментов и др.), и применяются нейронные сети для классификации сцен по этим признакам. Так, в [3] представлен подход детектирования драк, основанный на вычислении гистограммы направленных градиентов и гистограммы оптического потока. В [4] предлагается подход детектирования насильственного поведения в толпе, основанный на анализе изменения текстурных признаков Gray Level Co-occurrence Matrix (GLCM) во времени, в [5] – подход детектирования драк на основе анализа траекторий движения областей кадров, в [6] – применение сверточной нейронной сети для детектирования огнестрельного оружия, в [7] – подход детектирования ударов, основанный на оценке двумерного скелета человека. В работах [8, 9] предлагается применение сверточных и рекуррентных нейронных сетей, обучаемых от начала до конца, для детектирования насильственного поведения людей по видеозаписям.

Но для сложных протяженных во времени сцен использование нейронных сетей для конечной классификации не представляется возможным в силу отсутствия явных отличительных признаков и малого количества примеров для обучения сетей. Например, в одном из самых многочисленных наборов сцен правонарушений, UCF Crime [10], присутствует всего 9 примеров ограбления прохожих злоумышленниками на средстве передвижения, на котором злоумышленники подъезжают к жертве, обыскивают карманы, забирают ценности и уезжают без борьбы и нанесения ударов. Для распознавания таких сцен необходимо применение знаний экспертов об интересующих типах девиантного поведения, представленных в виде формальных моделей. Так, при детектировании бесхозных предметов в некоторых работах жестко кодируется модель, при которой предмет считается бесхозным, если он был статичен в течение определенного времени и рядом с ним отсутствовали люди [11]. В [12, 13] предлагается более сложная модель, призванная отслеживать возможного владельца предмета. В работе [14] предлагается формирование модели на языке Prolog для детектирования простой сцены ограбления, при которой двое или более людей встречаются, стоят вместе, и через некоторое время хотя бы один человек убегает. Но существующие работы ограничены применением простых малорепрезентативных признаков, таких как двумерная форма и скорость объектов с классификацией и отслеживанием объектов по этим же признакам, что ограничивает спектр распознаваемых сцен девиантного поведения. В то же время развитие технологий компьютерного зрения делает возможным оценку более тонких семантических признаков сцены по видеозаписям, таких как положение трехмерного скелета человека и элементарные действия людей (удары, падения, передача предмета и т. д.), и применение современных технологий компьютерного зрения может позволить распознавание куда более широкого спектра сцен девиантного поведения.

В настоящей работе предлагается метод автоматического распознавания протяженных во времени сцен девиантного поведения людей для поддержки принятия решений операторами систем видеомониторинга, интегрирующий технологии управления знаниями и современные технологии компьютерного зрения для детектирования и сегментации объектов, оценки трехмерного скелета человека, отслеживания объектов на кадрах видеозаписи, оценки нормали к поверхности земли и классификации элементарных действий людей по трехмерному скелету.

Статья состоит из четырех разделов. В первом разделе представлена постановка задачи автоматического распознавания протяженных во времени сцен девиантного поведения людей, во втором разделе – метод автоматического распознавания протяженных во времени сцен девиантного поведения, в третьем разделе – пример модели девиантного поведения, в четвертом разделе – результаты экспериментальных исследований предложенного метода на примере задачи обнаружения бесхозных предметов.

## 1. ПОСТАНОВКА ЗАДАЧИ

Обнаружение и противодействие девиантному поведению людей осуществляется посредством взаимодействия между системами умного города (системой видеонаблюдения, системой локального оповещения), службой видеомониторинга и службами реагирования (локальной службой охраны, полицией, скорой помощью и другими). Система поддержки принятия решений операторами при детектировании девиантного поведения должна обрабатывать исходные видеопотоки городских камер видеонаблюдения, осуществлять фильтрацию и при обнаружении инцидента предоставлять операторам информацию, которая позволяет оперативно принять решение о дальнейших действиях – скоординировать локальную службу охраны, полицию, скорую помощь, воспользоваться локальной системой голосового оповещения либо проигнорировать инцидент как ложный (рис. 1). Эта информация должна включать в себя тип предполагаемого девиантного поведения, отрывок видеозаписи с обозначенными участниками и объектами, а также обоснование выбора данного отрывка в виде обнаруженных фактов для заострения внимания оператора.

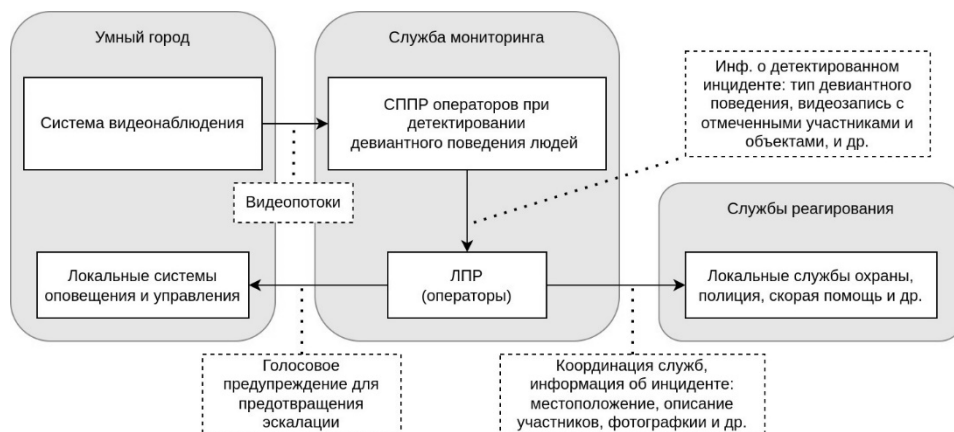


Рис. 1. Модель организации обнаружения и противодействия девиантному поведению людей при использовании СППР

Fig. 1. An organization model for detecting and counteracting deviant behavior of people when using DSS

Многие типы девиантного поведения не имеют явных отличительных признаков, на детектирование которых могла бы опираться СППР, и проявляются во взаимодействии между людьми и объектами в течение потенциально длительного периода времени. Так, при обнаружении группы людей

может быть непонятно, что наблюдается ограбление, но если обратить внимание на то, что злоумышленники подошли и остановили жертву и жертва передала злоумышленникам какой-то предмет, можно предположить, что происходит девиантное поведение. Или же если был замечен человек, наклонившийся у окна машины на парковке, можно предположить, что это его машина, но, если человек подходит к разным машинам и заглядывает в окна, можно предположить, что назревает ограбление. В данной работе ставится задача разработки метода автоматического распознавания девиантного поведения людей для поддержки принятия решений операторов систем видеомониторинга, ориентированного на протяженные во времени типы девиантного поведения.

## **2. МЕТОД АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ ПРОТЯЖЕННЫХ ВО ВРЕМЕНИ ТИПОВ ДЕВИАНТНОГО ПОВЕДЕНИЯ ЛЮДЕЙ**

Предлагаемый в рамках настоящей работы метод автоматического распознавания девиантного поведения людей подразумевает интеграцию различных технологий компьютерного зрения и управления знаниями в единой системе, концептуальная модель которой представлена на рис. 2. Система состоит из трех слоев: слой машинного восприятия, слой агрегации числовых признаков и формирования знаний, слой логического вывода.

Слой машинного восприятия осуществляет извлечение сырых признаков из кадров видеозаписи при помощи следующих технологий компьютерного зрения:

- детектирования и сегментации объектов (в том числе людей) на RGB-изображении [15];
- отслеживания (сопоставления) объектов между разными кадрами видеозаписи [16];
- оценки трехмерного скелета человека по RGB-изображению [17];
- оценки расстояния до объектов по RGB-изображению на основе детектированного вектора нормали к поверхности земли и заданных параметров видеокамеры [18];
- классификации действий людей по последовательности их трехмерного скелета [19].

Данный набор технологий позволяет определять относительное положение людей и объектов с течением времени, а также принимаемые позы, выполняемые действия и взаимодействие людей в виде следующих числовых признаков:

- классы объектов (человек, автомобиль, велосипед, мотоцикл, сумка);
- двумерные координаты обрамляющих рамок объектов на изображении;
- двумерные маски сегментации объектов на изображении;
- соответствие одних и тех же объектов на разных кадрах в виде глобальных идентификаторов;
- координаты ключевых точек скелета человека в трехмерном пространстве относительно центральной точки скелета;
- координаты объектов, соприкасающихся с плоскостью земли, в общем трехмерном пространстве;

- вероятности выполнения определенных кратковременных действий человеком или парой взаимодействующих людей: передача предмета, перетягивание предмета, обыск карманов, поза стрельбы из пистолета и др.

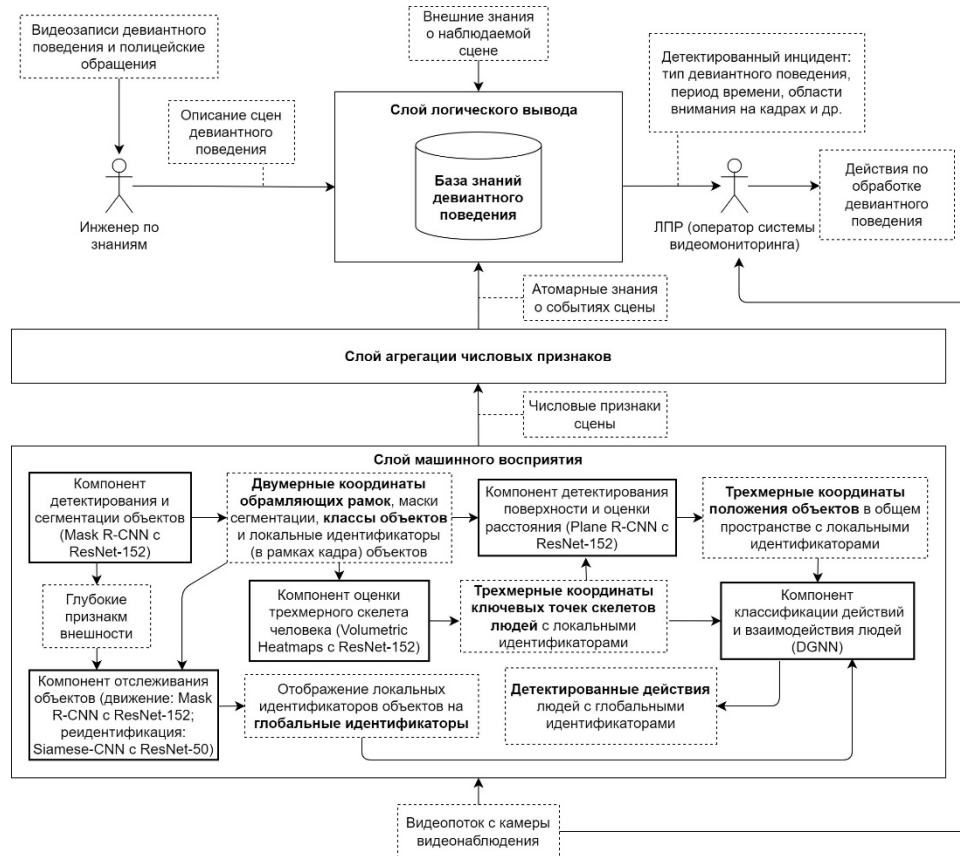


Рис. 2. Концептуальная модель СППР при детектировании девиантного поведения людей

Fig. 2. A conceptual model of DSS for detecting deviant behavior of people

Далее полученные признаки преобразуются в атомарные знания в слое агрегации числовых признаков (рис. 3) и передаются выше на слой логического вывода. Преобразование осуществляется посредством программной обработки, не требующей применения интеллектуальных технологий. Знания представляются фактами на языке Prolog, с которыми ассоциируются идентификаторы объектов (ObjId, Pid, VId, Bid), время кадра, из которого они были получены (Time), и другие атрибуты.

В рамках настоящей работы предлагается 7 типов атомарных знаний, на которых базируются модели типов девиантного поведения людей.

1. Двумерные обрамляющие рамки людей и объектов (bounds). На их основе можно сделать предположение о взаимодействии между людьми и объектами: находится ли человек на средстве передвижения, держит ли человек сумку и т. д.

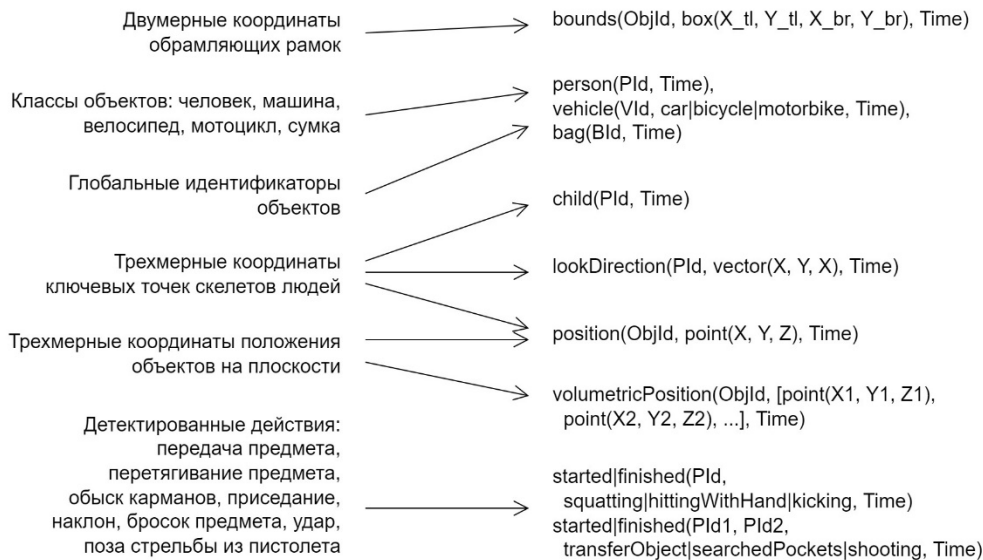


Рис. 3. Преобразование числовых признаков в факты на языке Prolog

Fig. 3. Converting numeric features into Prolog facts

2. Классы (person / vehicle / bag). Необходимы при описании взаимодействия людей и объектов.

3. Является ли человек маленьким ребенком (child). Может быть использован для обнаружения ситуации, при которой ребенок находится в потенциальной опасности без присмотра взрослого.

4. Трехмерный вектор направления головы человека (lookDirection). Может быть использован для определения взаимодействия между людьми, между человеком и объектом, а также для определения вида перемещения человека: вперед, боком или спиной.

5. Положение человека или объекта в общем трехмерном пространстве (position). Необходимо для определения перемещения, а также возможного взаимодействия между близко расположенными людьми и объектами.

6. Положение объектов в общем трехмерном пространстве в виде фигуры (volumetricPosotion). Может быть использовано для более точного определения взаимодействия между людьми и объектами.

7. Факты начала/завершения выполнения кратковременного действия (started/finished). Необходимы для определения состояния человека, а также типа взаимодействия между людьми и объектами.

Так, рассмотрев вектор направления головы, положение человека и машины на плоскости земли и распознав действие наклона, можно детектировать, что человек заглядывает в окно автомобиля. Если человек повторяет это действие с множеством разных машин на парковке, можно предположить намерение ограбления. Рассмотрев вектор направления головы, изменение положения человека на плоскости и позы человека (поза стрельбы из пистолета), можно предположить, что происходит перестрелка, при которой человек перемещается боком или спиной в позе стрельбы, не сводя глаз с противника.

В слое логического вывода происходит распознавание сцен девиантного поведения. Для этого с заданной периодичностью осуществляется логический вывод по базе знаний (БЗ), которая содержит три типа знаний:

1) описания типов девиантного поведения в виде логических правил, которые формируются инженером по знаниям на основе видеозаписей и текстовых описаний произошедших инцидентов девиантного поведения;

2) автоматически детектируемые динамические атомарные знания о наблюдаемых на видеозаписи событиях;

3) внешние статические и динамические атомарные знания, которые вносятся вручную или импортируются из внешних источников для каждой камеры. Например, место наблюдения (подземный переход, улица, парк и т. д.), время суток, текущая погода (сильный дождь, снег, жаркая погода и т. д.).

С динамическими знаниями ассоциируется время, которое соответствует времени происходящих на видеозаписи событий. Например, время начала и завершения определенного выполняемого человеком действия или время, соответствующее определенному положению человека в трехмерном пространстве. Новые динамические знания добавляются в БЗ с каждым кадром видеозаписи, при этом старые знания удаляются, если они не относятся к актуальному периоду времени. Размер актуального периода времени является настраиваемым параметром системы.

Материалы по распознанным инцидентам девиантного поведения передаются операторам системы видеомониторинга для принятия решений по их обработке. Материалы содержат видеозапись с отмеченными областями, на которые следует обратить внимание, и пояснения, на основе каких фактов СППР приняла решение о том, что данная сцена может содержать девиантное поведение. В случае, если девиантное поведение подтвердилось, оператор может связаться с локальной службой безопасности, полицией или скорой помощью для оказания помощи жертвам и задержания подозреваемых, а также воспользоваться локальной системой оповещения для предотвращения эскалации и оказания поддержки жертвам.

Достоинством предложенного метода является разбиение задачи распознавания специфичных сцен девиантного поведения на общие базовые подзадачи, для решения которых возможно создание достаточно объемных обучающих наборов данных моделей компьютерного зрения. Кроме того, в случае обнаружения ошибки распознавания целевой сцены причина ошибки легко локализуется и во многих случаях может быть устранена посредством корректировки описания сцены девиантного поведения или дополнения обучающего набора данных проблемной модели.

### **3. ПРИМЕР МОДЕЛИ ДЕВИАНТНОГО ПОВЕДЕНИЯ ЛЮДЕЙ**

В рамках предложенного метода при распознавании сцен девиантного поведения СППР выступает в роле фреймворка, принимающего модели девиантного поведения людей в качестве конфигурации. Рассмотрим в качестве примера модель размещения бесхозной сумки в публичном месте. Во многих существующих системах используются довольно простые критерии детектирования бесхозной сумки, которые могут привести к ложным срабатываниям



в случае специфичного поведения людей. Например, в системах Macroscop<sup>6</sup> и Starvision<sup>7</sup> детектируется предмет, который был оставлен без движения в течение длительного периода времени, но это условие не учитывает ситуацию, при которой владелец находится поблизости в ожидании чего-то. В некоторых работах присутствуют попытки отслеживания возможного владельца или владельцев сумки среди людей, находившихся поблизости [12, 13]. В настоящей работе предлагается дополнительное условие – **отсутствие наблюдателя**: сумка считается бесхозной, если среди тех людей, кто находился рядом с сумкой в момент ее размещения, нет человека, которые не отошел бы от сумки надолго, не оставив около нее наблюдателя.

Модель размещения бесхозной сумки представлена на рис. 4 в нотации, основанной на языке Prolog. Для детектирования бесхозной сумки (`devBeh(abandonedBag(Bag))`) определяется период времени, когда сумка была размещена (`wasPlaced(Bag)`). Далее проверяется, что среди всех возможных владельцев сумки, находившихся поблизости в период размещения, нет человека, который бы не отошел от нее более чем на 30 секунд без наблюдателя (`possibleOwnerWhoDidntLeaveAPlacedBagForMoreThan30SecWithoutAWatcher(...)`). Поскольку определение положения сумки в трехмерном пространстве осуществляется с предположением, что она была размещена на земле, необходимо исключить ситуацию, при которой человек держит сумку над землей и стоит неподвижно. Для этого проверяется аналогичное условие, но в качестве отношения близости рассматривается пересечение обрамляющих рамок сумки и человека (`possibleOwnerWhoWasntSeparatedFromAPlacedBagForMoreThan30SecWithoutAWatcher(...)`). Помимо размещения сумки, исходным моментом также является обнаружение уже оставленной сумки (`stillBagDiscovered(Bag)`). В этом случае в качестве возможного владельца мы рассматриваем всех людей поблизости в последующие 30 секунд, предполагая, что человек мог ненадолго отойти без наблюдателя.

Считается, что сумка была размещена (`wasPlaced(Bag)`), если за прошедшие 5 секунд произошло существенное смещение ее обрамляющей рамки (`moved(Bag)`), и оно не происходило в последующие 5 секунд, при этом сумка присутствовала в кадре (`present(Bag)`). Существенное смещение обрамляющей рамки (`moved(Bag)`) произошло, если площадь пересечения рамки за два рассматриваемых момента времени равна менее половины общей площади рамок (`intersectLessThan(Box1, Box2, 0.5)`). Считается, что сумка присутствовала в кадре (`present(Bag)`), если она была видна в кадре либо если она была перекрыта в течение менее 15 секунд и не сместилась существенно за это время. Считается, что уже размещенная сумка была обнаружена (`stillBagDiscovered(Bag)`), если существенного смещения обрамляющей рамки не наблюдалось в течение 5 секунд после ее появления.

---

<sup>6</sup> Детектор оставленных предметов. Доступен по: <https://macroscop.com/produkty/programma-dlya-ip-kamer/detektor-ostavlennyh-predmetov> (дата обращения: 17.07.2022).

<sup>7</sup> Система обнаружения оставленных предметов. Доступен по: <https://satvision-cctv.ru/upload/iblock/d13/4.-SISTEMA-OBNARUZHENIYA-OSTAVLENNYKH-PREDMETOV.pdf> (дата обращения: 17.07.2022).

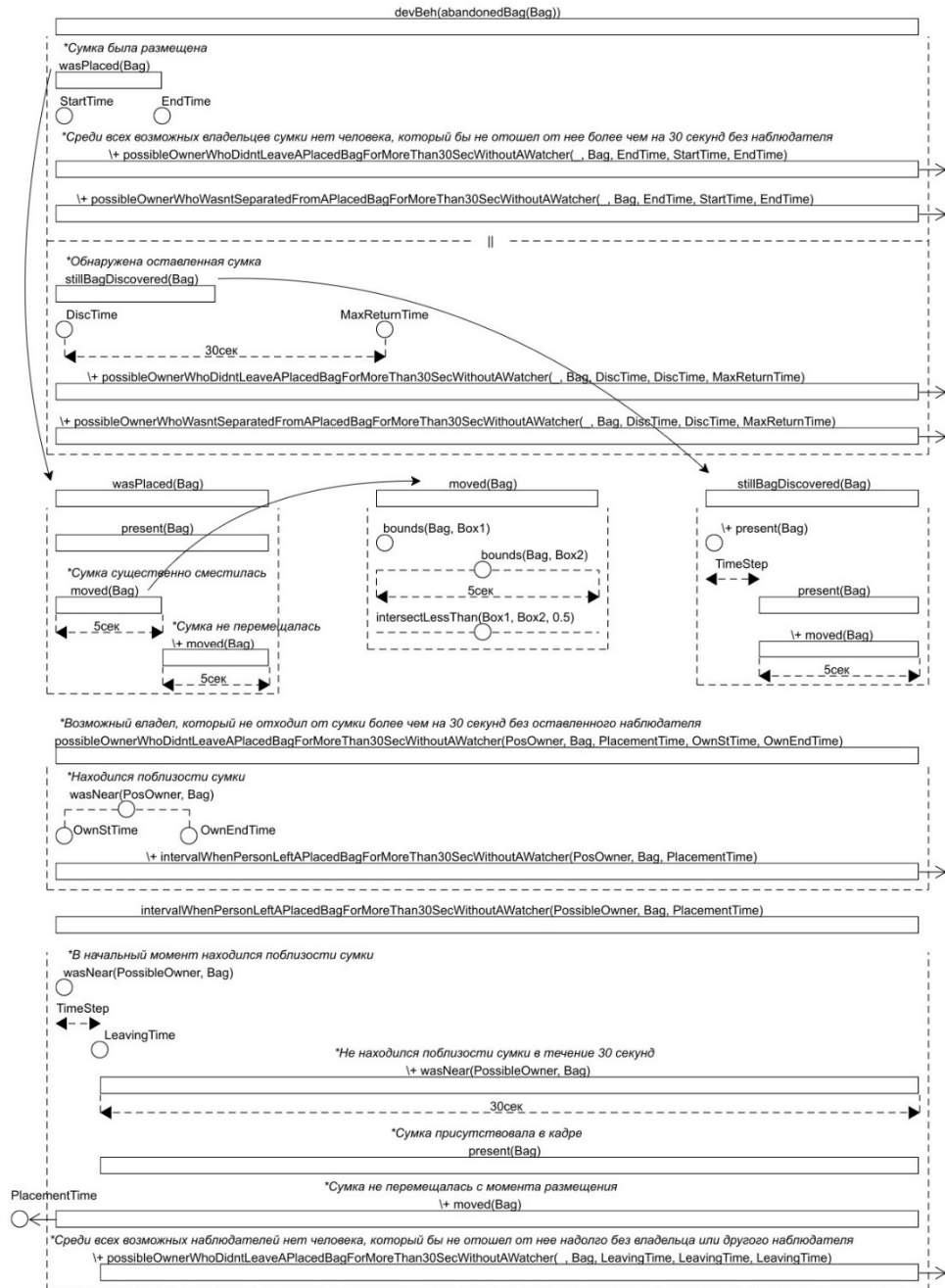


Рис. 4. Модель размещения бесхозной сумки в публичном месте, представленная в нотации на основе языка Prolog

Fig. 4. A model for placing an abandoned bag in a public place presented in Prolog-based notation

Возможный владелец сумки, который не отходил от нее более чем на 30 секунд, не оставив после себя наблюдателя (`possibleOwnerWhoDidntLeaveAPlacedBagForMoreThan30SecWithoutAWatcher(...)`), определяется среди тех людей, кто находился поблизости сумки

(wasNear(PossibleOwner, Bag)) в определенный период времени, зависящий от контекста. Считается, что человек находится поблизости сумки, если расстояние между ним и сумкой не превышает трех метров. Интервал, в котором человек отходил от сумки более чем на 30 секунд, не оставив наблюдателя (intervalWhenPersonLeftAPlacedBagForMoreThan30SecWithoutAWatcher(...)), детектируется от момента размещения сумки. Определяется период времени длиной в 30 секунд, когда сумка присутствовала в кадре, но человек не находился поблизости, и сумка не перемещалась существенно с момента размещения. Кроме того, рекурсивно проверяется, что среди всех людей, находившихся поблизости сумки в момент ухода владельца, не было человека – возможного наблюдателя, который также не отходил бы от сумки надолго без другого наблюдателя или возвращения владельца.

#### **4. ТЕСТИРОВАНИЕ МЕТОДА АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ ПРОТЯЖЕННЫХ ТИПОВ ДЕВИАНТНОГО ПОВЕДЕНИЯ ЛЮДЕЙ НА ПРИМЕРЕ РАЗМЕЩЕНИЯ СУМКИ В ПУБЛИЧНОМ МЕСТЕ**

Для тестирования предложенного метода на его основе был разработан прототип системы автоматического распознавания протяженных типов девиантного поведения. Для решения отдельных задач компьютерного зрения использовались открытые модели нейронных сетей, представленные в работах [15–19]. Полученный прототип и модель размещения сумки в публичных местах были протестированы на наборе данных PETS-2006 [20]. Набор содержит 7 сцен, снятых четырьмя различными камерами видеонаблюдения с разрешением  $768 \times 576$  пикселей. Для тестирования использовались видеозаписи, снятые тремя камерами, а четвертая, как и во многих работах, была исключена в силу сильной удаленности от детектируемых событий. В наборе присутствует три типа сценариев:

- 1) человек оставил сумку и ушел через какое-то время (рис. 5);
- 2) человек положил сумку на землю, но через какое-то время забрал ее (рис. 6);
- 3) человек оставил сумку с другим человеком – наблюдателем (рис. 7).

При этом интерпретация 3-го типа сценария, в котором владелец оставляет сумку с другим человеком, остается открытой, и во многих работах также считается проблемным инцидентом. Но в рамках настоящей работы мы считаем, что эта ситуация не является инцидентом, на который необходимо реагировать. Данная интерпретация позволяет продемонстрировать возможность учета специфичных случаев взаимодействия людей при применении предложенного метода.

В результате тестирования были корректно распознаны примеры всех трех типов сценариев, но в шести из двадцати одной видеозаписи использованная для детектирования и сегментации объектов нейронная сеть [15] не смогла детектировать сумку, из-за чего сцена не распозналаась. Причиной ошибки стала специфичная форма и углы обзора сумки. Для устранения данной ошибки возможным решением является дополнение обучающего набора модели детектирования объектов проблемными примерами.



Рис. 5. Пример размещения сумки в публичном месте из набора PETS-2006

Fig. 5. An example of placing a bag in a public place from the PETS-2006 dataset

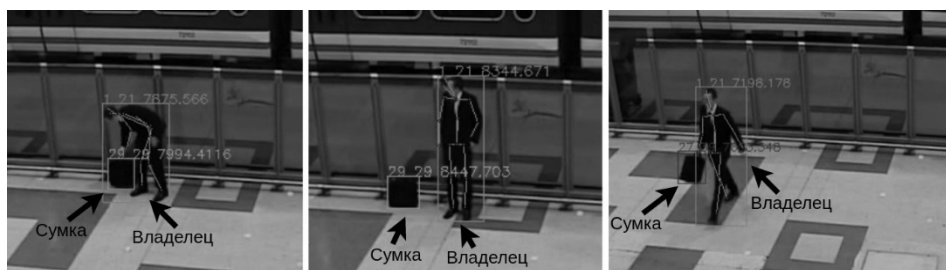


Рис. 6. Пример временного размещения сумки в публичном месте из набора PETS-2006.

Fig. 6. An example of temporary placement of a bag in a public place from the PETS-2006 dataset

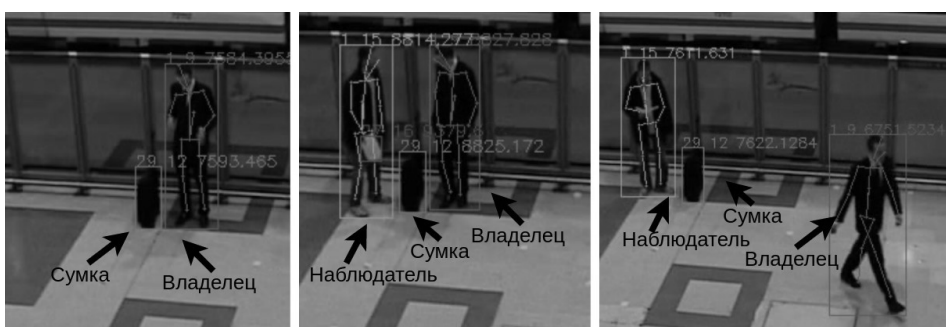


Рис. 7. Пример оставления сумки в публичном месте с наблюдателем из набора PETS-2006

Fig. 7. An example of leaving a bag in a public place with a watcher from the PETS-2006 dataset

## ЗАКЛЮЧЕНИЕ

В рамках настоящей работы был представлен метод автоматического распознавания протяженных во времени сцен девиантного поведения людей для поддержки принятия решений операторами систем видеомониторинга, отличающийся интеграцией современных технологий компьютерного зрения и технологии управления знаниями в единой системе. Данная система является фреймворком, принимающим в качестве конфигураций модели интересующих типов девиантного поведения. В качестве примера модели протяженного типа девиантного поведения была представлена модель сцены размещения сумки в публичном месте, отличающаяся учетом возможного появления наблюдателя при уходе владельца. На основе полученного метода был разработан прототип системы распознавания протяженных типов девиантного поведения с использованием открытых моделей нейронных сетей для решения отдельных задач компьютерного зрения. Полученный прототип и модель размещения сумки в публичном месте были протестированы на наборе данных PETS-2006. В результате тестирования прототип смог распознать все три типа сценариев, представленных в наборе данных, но имел ложноотрицательные срабатывания из-за ошибок детектирования сумки нейронной сетью для детектирования и сегментации объектов. Направлениями дальнейших исследований являются тестирование разработанного метода на других типах девиантного поведения и повышение точности применяемых моделей компьютерного зрения. Результаты настоящей работы могут быть использованы для разработки СППР операторов систем видеомониторинга при детектировании девиантного поведения людей, позволяющей оперативное обнаружение и принятие своевременных мер по разрешению инцидентов и оказанию помощи жертвам.

## СПИСОК ЛИТЕРАТУРЫ

1. Valikhujaev Y., Abdusalomov A., Cho Y.I. Automatic fire and smoke detection method for surveillance systems based on dilated CNNs // *Atmosphere*. – 2020. – Vol. 11, N 11. – P. 1241. – DOI: 10.3390/atmos11111241.
2. Computer vision-based accident detection in traffic surveillance / E.P. Ijjina, D. Chand, S. Gupta, K. Goutham // 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). – Kanpur, India, 2019. – P. 1–6. – DOI: 10.1109/ICCCNT45670.2019.8944469.
3. Violence detection in video using computer vision techniques / E.B. Nievas, O.D. Suarez, G.B. García, R. Sukthankar // *Computer Analysis of Images and Patterns, CAIP 2011*. – Berlin: Springer, 2011. – P. 332–339. – DOI: 10.1007/978-3-642-23678-5\_39.
4. Lloyd K., Rosin P., Marshall D., Moore S. Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures // *Machine Vision and Applications*. – 2017. – Vol. 28. – P. 361–371.
5. Spatio-temporal elastic cuboid trajectories for efficient fight recognition using Hough forests / I. Serrano, O. Deniz, G. Bueno, G. Garcia-Hernando, T.-K. Kim // *Machine Vision and Applications*. – 2018. – Vol. 29. – P. 207–217.
6. Weapon detection using YOLO V3 for Smart Surveillance System / S. Narejo, B. Pandey, D. Esenarro Vargas, C. Rodriguez, M.R. Anjum // *Mathematical Problems in Engineering*. – 2021. – DOI: 10.1155/2021/9975700.
7. Singh A., Patil D., Omkar S.N. Eye in the Sky: real-time drone surveillance system (DSS) for violent individuals identification using ScatterNet Hybrid Deep Learning Network // 2018 IEEE/CVF

Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). – Salt Lake City, UT, USA, 2018. – P. 1710–1718. – DOI: 10.1109/CVPRW.2018.00214.

8. Violent Interaction Detection in Video Based on Deep Learning / P. Zhou, Q. Ding, H. Luo, X. Hou // *Journal of Physics: Conference Series*. – 2017. – Vol. 844. – P. 012044.

9. Sudhakaran S., Lanz O. Learning to detect violent videos using convolutional long short-term memory // 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). – Lecce, Italy, 2017. – Vol. 1. – P. 1–6. – DOI: 10.1109/AVSS.2017.8078468.

10. Sultani W., Chen C., Shah M. Real-world anomaly detection in surveillance videos // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. – Salt Lake City, UT, USA, 2018. – P. 6479–6488. – DOI: 10.1109/CVPR.2018.00678.

11. Liao H., Chang J., Chen L. A localized approach to abandoned luggage detection with foreground-mask sampling // 2008 IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance. – Santa Fe, NM, USA, 2008. – P. 132–139. – DOI: 10.1109/AVSS.2008.9.

12. Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance / K. Lin, S. Chen, C. Chen, D. Lin, Y. Hung // *IEEE Transactions on Information Forensics and Security*. – 2015. – Vol. 10, N 7. – P. 1359–1370.

13. Park H., Park S., Joo Y. Robust detection of abandoned object for smart video surveillance in illumination changes // *Sensors*. – 2019. – Vol. 19, N 23. – P. 5114.

14. Morozov A.A., Sushkova O. Real-Time analysis of video by means of the Actor Prolog language // *Computer Optics*. – 2018. – Vol. 40. – P. 947–957.

15. He K., Girshick R., Dollar P. Rethinking ImageNet Pre-Training // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). – Seoul, Korea (South), 2019. – P. 4917–4926. – DOI: 10.1109/ICCV.2019.00502.

16. Bergmann P., Meinhardt T., Leal-Taixe L. Tracking without Bells and Whistles // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). – Seoul, Korea (South), 2019. – P. 941–951. – DOI: 10.1109/ICCV.2019.00103.

17. Sun X., Li C., Lin S. An integral pose regression system for the ECCV2018 PoseTrack Challenge // *arXiv preprint arXiv:1809.06079*.

18. Ryabchikov I., Teslya N. Estimating position of multiple people in common 3D space via city surveillance cameras // 2021 28<sup>th</sup> Conference of Open Innovations Association (FRUCT). – Moscow, Russia, 2021. – P. 390–397. – DOI: 10.23919/FRUCT50888.2021.9347579.

19. Skeleton-based action recognition with directed graph neural networks / L. Shi, Y. Zhang, J. Cheng, H. Lu // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). – Seoul, Korea (South), 2019. – P. 7904–7913.

20. PETS 2006. – URL: <https://mldta.com/dataset/pets-2006/> (accessed: 17.07.2022).

*Рябчиков Игорь Александрович*, младший научный сотрудник лаборатории интегрированных систем автоматизации Санкт-Петербургского федерального исследовательского центра Российской академии наук. Основное направление научных исследований – интеллектуальные системы и технологии, распределенные вычисления. Имеет более 10 печатных работ. E-mail: i.a.ryabchikov@gmail.com

*Ryabchikov Igor A.*, junior researcher in the laboratory of computer-aided integrated systems of the St. Petersburg Federal Research Center of the Russian Academy of Sciences. His research interests are currently focused on intelligent systems and technologies, as well as distributed computing. He has more than 10 publications. E-mail: i.a.ryabchikov@gmail.com

***A method for automatic recognition of deviant behavior of people based on the integration of computer vision and knowledge management technologies to support decision-making by operators of video monitoring systems\****

**I.A. RYABCHIKOV**

*SPC RAS, 39 14-ya Liniya Vasil'yevskogo Ostrova, St. Petersburg, 199178, Russian Federation  
i.a.ryabchikov@gmail.com*

**Abstract**

Ensuring safety of people in the urban environment is an important area for improving the quality of people's lives, and the development of modern intelligent technologies creates new opportunities to achieve this goal. The use of modern intelligent technologies can realize the potential of video surveillance systems, making possible the automatic recognition of dangerous situations in real time in order to take timely measures to handle them and provide aid to victims. Often, a dangerous situation is caused by deviant behavior of people – robbery, fight, vandalism, etc. But the existing works focused on recognizing deviant behavior are only focused on recognizing short-term distinguishing features, such as punches, falls or weapons in the hands of a person. At the same time, such features may often be absent, for instance, when a passerby is robbed but no fighting occurs, which is why the task of recognizing complex long-term scenes of deviant behavior remains unresolved. This paper proposes a method for automatic recognition of long-term human deviant behavior scenes, characterized by the integration of knowledge management and computer vision technologies for detecting and segmenting objects, estimating the three-dimensional human skeleton, tracking objects in video, estimating the ground plane normal to calculate the camera distance, and classification of short-term actions of people using three-dimensional skeleton. This method can be used in the development of a decision support system by operators of video monitoring systems used to detect and handle deviant behavior of people in real time in order to prevent escalation, provide timely aid to victims and detain suspects.

**Keywords:** knowledge base, abandoned object, deviant behavior, action classification, computer vision, robbery, decision support, scene recognition, video surveillance system, three-dimensional skeleton

**REFERENCES**

1. Valikhujaev Y., Abdusalomov A., Cho Y.I. Automatic fire and smoke detection method for surveillance systems based on dilated CNNs. *Atmosphere*, 2020, vol. 11, no. 11, p. 1241. DOI: 10.3390/atmos11111241.
2. Ijjina E.P., Chand D., Gupta S., Goutham K. Computer vision-based accident detection in traffic surveillance. *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, 2019, pp. 1–6. DOI: 10.1109/ICCCNT45670.2019.8944469.
3. Nieves E.B., Suarez O.D., García G.B., Sukthankar R. Violence detection in video using computer vision techniques. *Computer Analysis of Images and Patterns, CAIP 2011*. Berlin, Springer, 2011, pp. 332–339. DOI: 10.1007/978-3-642-23678-5\_39.
4. Lloyd. K., Rosin. P., Marshall. D., Moore S. Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures. *Machine Vision and Applications*, 2017, vol. 28, pp. 361–371.
5. Serrano I., Deniz O., Bueno G., Garcia-Hernando G., Kim T.-K. Spatio-temporal elastic cuboid trajectories for efficient fight recognition using Hough forests. *Machine Vision and Applications*, 2018, vol. 29, pp. 207–217.

---

\* Received 12 July 2022.

6. Narejo S., Pandey B., Esenarro vargas D., Rodriguez C., Anjum M.R. Weapon detection using YOLO V3 for Smart Surveillance System. *Mathematical Problems in Engineering*, 2021. DOI: 10.1155/2021/9975700.
7. Singh A., Patil D., Omkar S.N. Eye in the Sky: real-time drone surveillance system (DSS) for violent individuals identification using ScatterNet Hybrid Deep Learning Network. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, 2018, pp. 1710–1718. DOI: 10.1109/CVPRW.2018.00214.
8. Zhou P., Ding Q., Luo H., Hou X. Violent Interaction detection in video based on deep learning. *Journal of Physics: Conference Series*, 2017, vol. 844, p. 012044.
9. Sudhakaran S., Lanz O. Learning to detect violent videos using convolutional long short-term memory. *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Lecce, Italy, 2017, vol. 1, pp. 1–6. DOI: 10.1109/AVSS.2017.8078468.
10. Sultani W., Chen C., Shah M. Real-world anomaly detection in surveillance videos. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, 2018, pp. 6479–6488. DOI: 10.1109/CVPR.2018.00678.
11. Liao H., Chang J., Chen L. A localized approach to abandoned luggage detection with foreground-mask sampling. *2008 IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance*, Santa Fe, NM, USA, 2008, pp. 132–139. DOI: 10.1109/AVSS.2008.9.
12. Lin K., Chen S., Chen C., Lin D., Hung Y. Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance. *IEEE Transactions on Information Forensics and Security*, 2015, vol. 10, no. 7, pp. 1359–1370.
13. Park H., Park S., Joo Y. Robust detection of abandoned object for smart video surveillance in illumination changes. *Sensors*, 2019, vol. 19, no. 23, p. 5114.
14. Morozov A.A., Sushkova O. Real-Time analysis of video by means of the Actor Prolog language. *Computer Optics*, 2018, vol. 40, pp. 947–957.
15. He K., Girshick R., Dollar P. Rethinking ImageNet Pre-Training. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 4917–4926. DOI: 10.1109/ICCV.2019.00502.
16. Bergmann P., Meinhardt T., Leal-Taixe L. Tracking Without Bells and Whistles. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 941–951. DOI: 10.1109/ICCV.2019.00103.
17. Sun X., Li C., Lin S. An integral pose regression system for the ECCV2018 PoseTrack Challenge. *arXiv preprint arXiv:1809.06079*.
18. Ryabchikov I., Teslya N. Estimating position of multiple people in common 3D space via city surveillance cameras. *2021 28<sup>th</sup> Conference of Open Innovations Association (FRUCT)*, Moscow, Russia, 2021, pp. 390–397. DOI: 10.23919/FRUCT50888.2021.9347579.
19. Shi L., Zhang Y., Cheng J., Lu H. Skeleton-based action recognition with directed graph neural networks. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 7904–7913.
20. PETS 2006. Available at: <https://mldta.com/dataset/pets-2006/> (accessed 17.07.2022).

Для цитирования:

Рябчиков И.А. Метод автоматического распознавания девиантного поведения людей на основе интеграции технологий компьютерного зрения и управления знаниями для поддержки принятия решений операторами систем видеомониторинга // Системы анализа и обработки данных. – 2022. – № 3 (87). – С. 21–36. – DOI: 10.17212/2782-2001-2022-3-21-36.

For citation:

Ryabchikov I.A. Metod avtomaticheskogo ispol'zovaniya deviantnogo povedeniya lyudei na osnove samykh vysokikh tekhnologii komp'yuternogo zreniya i upravleniya trebovaniyami dlya podderzhki prinyatiya reshenii operatorami sistem videomonitoringa [A method for automatic recognition of deviant behavior of people based on the integration of computer vision and knowledge management technologies to support decision-making by operators of video monitoring systems]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2022, no. 3 (87), pp. 21–36. DOI: 10.17212/2782-2001-2022-3-21-36.