

## ОБЗОР

УДК 004.853

**ИНТЕЛЛЕКТУАЛЬНЫЕ МОБИЛЬНЫЕ РОБОТЫ  
И АНАЛИЗ ИХ ДЕЯТЕЛЬНОСТИ****Иванько А.Ф., Иванько М.А., Ибрагимов А.А.***Московский политехнический университет, Москва, e-mail: alekfed@mail.ru,  
mihaleks@mail.ru, tima.kurkov7@mail.ru*

Удача интеллектуальных мобильных роботов, трудящихся, а также действующих совместно с людьми, находится в зависимости с их возможностями подводить итог и исследовать человеческие движения, а также извлекать единое представление наблюдаемой картинке. В данной статье мы постараемся исследовать человеческие движения, исполняемые в реальных обстоятельствах с непрерывного исследования мобильного робота. Для данной задачи человеческая деятельность обуславливается как модифицирующаяся пластическая структура туловища человека, взаимодействующая с основными предметами, обеспечивающими в таком случае работоспособность в границах сферы обитания. Для того чтобы сократить ограничения мобильного робота, которые возникают из-за его неполноценных воспринимающих модальностей, возможно зрительные исследования отразить в абстрактное высококачественное пространство. Для того чтобы подытожить стандарты, нежелательные к четким численным позициям в нашем обществе. Несколько высококачественных объемно-временных репрезентаций применяются с целью регистрации разных нюансов взаимоотношений между человеком и его обществом. Подобно поиску данных в каком-либо месте, генеративно-вероятностный способ применяется с целью возобновления тайных, семантически важных определений в закодированных наблюдениях абсолютно разными способами. Небольшое число раскрытых определений смотрятся ровно как типы человеческой деятельности, то, что предоставляет роботу зрительно наблюдаемые непростые действия человека. В конечном итоге заключение применяется с целью облегчения градиционного и постоянного обновления подобных определений, что дает возможность мобильному роботу результативно исследовать и со временем освежать собственные личные модификации деятельности людей, что приводит к результативной учебе в течение его существования.

**Ключевые слова:** искусственный интеллект, мобильные роботы, обучение, анализ человеческой деятельности, постоянное обучение

**INTELLIGENT MOBILE ROBOTS AND ANALYSIS OF THEIR ACTIVITIES****Ivanko A.F., Ivanko M.A., Ibragimov A.A.***Moscow Polytechnic University, Moscow, e-mail: alekfed@mail.ru,  
mihaleks@mail.ru, tima.kurkov7@mail.ru*

The success of intelligent mobile robots, working as well as working in conjunction with humans, depends on their ability to sum up and explore human movements, as well as extract a single representation of the observed image. In this article, we will try to explore human movements performed in real circumstances with the continuous study of a mobile robot. For this task, human activity is conditioned as a modifying plastic structure of the human body, interacting with the main objects that provide in this case, efficiency within the boundaries of the habitat. In order to reduce the limitations of the mobile robot, which arise from its defective perceiving modalities, it is possible to reflect visual research in an abstract high-quality space. In order to sum up the standards, we need clear numerical positions in our society. Several high-quality space-time representations are used to record various nuances of the relationship between a person and his society. Like searching for data in a place, the generative-probabilistic method is used to renew secret, semantically important definitions in coded observations in completely different ways. A small number of open definitions look exactly like the types of human activity that provides the robot visually observed complex human actions. In the end, the conclusion is used to facilitate the gradation and constant updating of such definitions, which allows the mobile robot to effectively explore and eventually refresh their own personal modifications of human activities, which leads to effective learning during its existence.

**Keywords:** artificial intelligence, mobile robots, training, analysis of human activity, continuous training

Автономные мобильные робототехнические платформы требуют непрерывного обновления своих собственных знаний о мире, которые они могут получить, прежде всего, благодаря многочисленным наблюдениям и взаимодействиям с людьми [1; 2]. Естественно, для того чтобы робот был полезен людям, ему нужно обучаться на протяжении всей своей жизни с человеком. Всяческие системы обучения, которые не требуют вмешательства людей в процесс обучения, в течение такого длительного пе-

риода могут помочь мобильным роботам стать более полезными, особенно при проживании интеллектуальных машин вместе с людьми. Если убрать человека из процесса обучения, то есть полностью довериться самообучению робота, то они, в принципе, могут недорого и быстро учиться с помощью огромного количества доступных данных, это позволяет им адаптироваться к окружающей среде, а также экономить время и усилия людей. Поддержание понимания динамической среды обитания чело-

века, то есть того, какие виды деятельности присутствуют в жизни людей, а какие уже «отмирают», позволяет машине корректировать свое собственное поведение, а также, когда она уже научится увиденному, помогать в выполнении наблюдаемой задачи.

Цели исследования:

1. Определить качественную пространственно-временную и векторную пространственную структуру для кодирования действий человека, наблюдаемых автономным мобильным роботом.

2. Разобрать методы изучения представлений, общих и повторяющихся моделей из нескольких закодированных визуальных наблюдений робота за действиями человека с использованием неконтролируемых вероятностных методов.

3. Рассмотреть работу человека с автономными мобильными роботами, которые запоминают непрерывные видеозаписи действий человека в процессе обучения на протяжении всей своей жизни совместно с людьми.

#### Материалы и методы исследования

Методология основана, во-первых, на обнаружении и отслеживании движений человеческого тела с помощью встроенных датчиков мобильного робота, во-вторых, на изучении местоположения объектов, которыми пользуется человек, в окружающей среде с использованием готовых методов. Каждое наблюдение, первоначально записанное как последовательность количественных положений тела, кодируется с использованием множественных качественных исчислений для преобразования точных пространственных и временных деталей наблюдения и, наконец, представляется в виде вектора появления дискретных качественных дескрипторов. Несколько скрытых тем извлекаются из наблюдений и рассматриваются как классы человеческой деятельности, каждый из которых определяется как полиномиальное распределение по автоматически сгенерированному «словарю» действий [3]. Представлены две методики для изучения представлений о человеческой деятельности. Первый – это вероятностный подход к аппроксимации низкого ранга, плохо работает с предварительно сегментированными видеорядами наблюдаемой человеческой деятельности. Второй – это более сложный вероятностный метод скрытого распределения Дирихле [4], несколько снимает требование ручной временной сегментации записанных наблюдений, позволяя роботу получить доступ к большим количествам данных, которые в противном случае потребовали

бы человеческого вмешательства. Метод Дирихле – это иерархическая модель, в которой каждое наблюдение моделируется как смесь по базовому набору тем, а каждая тема, в свою очередь, моделируется как смесь по дискретному словарю. Насколько известно, ранее этот подход не использовался, вероятностный подход, такой как метод Дирихле, с качественным пространственным представлением для восстановления представлений человеческой деятельности, наблюдаемых мобильным роботом. Эта работа переходит от использования стандартного набора данных, где каждый отдельный образец данных состоит из временно разделенного экземпляра отдельного действия человека, к более реалистичной настройке, где экземпляры расположены в более длинной последовательности наблюдений. Это совершенно спокойно расценивается как отрицание предположения о том, что люди непрерывно выполняют последовательность интересных действий, когда они находятся в поле зрения мобильного робота. Более разумное предположение состоит в том, что наблюдение за человеком моделируется как вероятностная смесь по основному числу скрытых тем, где некоторые темы можно считать «интересными» человеческими действиями.

Анализ человеческой деятельности с помощью мобильных роботов является недавней областью исследований, отчасти благодаря достижениям в области навигации, локализации и планирования с использованием вероятностных методов робототехники [5]. Это позволило мобильным роботам иметь более точные и надежные оценки своего собственного местоположения в окружающей среде и лучше выполнять действия, основанные на этих оценках.

В литературе существует общее различие между анализом человеческой деятельности на основе зрения, которое извлекает информацию из видеокamer с использованием методов компьютерного зрения, и сенсорами, носимыми вычислительными системами [6; 7]. Сенсорные системы часто полагаются на наличие небольших датчиков, а именно носимых датчиков, смартфонов или объектов с радиочастотной идентификацией (RFID), которые могут быть прикреплены к наблюдаемому человеку, чтобы получить представление о движениях этого человека. Мы, прежде всего, изучаем человеческую деятельность из визуальных данных, где понятие наблюдения ограничено полем зрения одной камеры, закрепленной на роботе. Это достаточно широкое подполе искусственного интеллекта, и читатель указывает на обзорные ста-

ты, которые подробно освещают эту тему, используя в основном статические камеры RGB [8] или 3D-камеры глубины [9; 10]. Тем не менее многие из распространенных методов в этих вопросах выполняют контролируемое обучение, где каждая обучающая ступень требует ручной сегментации и аннотирования с меткой основной истины. Это не является возможным решением для мобильного робота, который в идеале имеет минимальный контроль со стороны человека при использовании в реальном мире [11]. Методы неконтролируемого обучения считаются более подходящими для этой задачи, поскольку они не занимают много времени, вмешательства людей [12].

Многие работы в этой области использовали латентный семантический анализ (LSA) [13], вероятностный LSA и метод Дирихле для изучения категорий человеческой деятельности в неконтролируемой обстановке. Ученые объединили эти методы с низкоуровневыми пространственно-временными функциями точки интереса (STIP) для изучения категорий действий; локальные дескрипторы контекста формы на силуэтных изображениях [14]; сочетание семантических и структурных признаков для изучения действий, лиц и жестов рук; и путем слияния словаря локальных пространственно-временных объемов (кубоидов) со словарем спин-образов для фиксации деформации формы актера. Однако основной проблемой, упоминаемой в этих работах, является отсутствие пространственной информации о человеческом теле, захваченной низкокачественными изображениями, и отсутствие более долгосрочной временной информации, закодированной в признаках. Описательные пространственно-временные коррелограммы были использованы ранее для решения этой проблемы, однако этот подход все еще страдает от слабости обработки изображений, низкого уровня присвоения к человеку, который должен присутствовать в сцене в течение определенного времени обучения. Другой подход состоял в том, чтобы изучить временные отношения между действиями человека в неконтролируемой обстановке, чтобы точно представить человеческую деятельность [15]. Однако входные видео для этой техники требуют ручной временной сегментации перекрывающихся клипов фиксированной длины, что делает данную технику непомерно дорогой для обучения в течение всей жизни мобильного робота. Кроме того, каждая из этих техник была выполнена без вариативности системы отчета мобильного робота и ограничена обучением на временно сегментированных видеоданных во время

фазы обучения. Для решения этих проблем наблюдаемые действия человека и объекты преобразуются в качественное пространственное представление. Есть некоторые данные, позволяющие предположить, что существуют выделенные области мозга для выполнения таких преобразований. Поэтому, естественно, нужно попытаться встроить это в систему, чтобы понять поведение человека в видеоданных и в конечном счете встроить в роботизированные системы, чтобы представить поведение, выполняемое в среде, в которой они обитают. Существует много типов, выделенных в литературе. Некоторые из наиболее популярных типов: топологические, направленные и не топологические [16]. Обзор популярных типов приведен в качественные пространственные представления, которые часто используются для представления визуальных, количественных данных, наблюдений в малоразмерном и семантически более значимом качественном пространстве. Часто выполняется объектно-ориентированное преобразование видеопоследовательности, затем изучаются общие расположения преобразованных объектов с использованием различных ходов, например общие настройки места стола для приема пищи; простые действия для повседневной жизни из набора данных статической камеры [17]; удаление несогласованных визуальных наблюдений из видеопоследовательностей; и даже выполнение рассуждений о наблюдаемых пространственно-временных событиях [18]. Каждый из этих методов выполнялся в условиях контролируемого обучения, в отличие от упомянутых неконтролируемых методов. Кроме того, они широко используются для задач более глубокого рассуждения и применяются ко многим реальным областям, однако это выходит за рамки данной работы. Качественные пространственные положения могут быть заранее заданы вручную либо обнаружены на основе постоянно обновляемых наблюдений. Преимущество обучения положениям автоматически заключается в том, что они мгновенно становятся релевантными поведению наблюдаемой области. Однако общим ограничением является то, что часто все данные (или некоторая их выборка) должны существовать до того, как начинается обучение, в отличие от определения положения заранее.

Были изучены сложные пространственно-временные отношения между отслеживаемыми областями, представляющими движущиеся объекты в реальных областях жизни, например транспортные средства на участке автомагистрали, представленные в виде неких траекторий. Топологические

соотношения были также изучены путем создания относительного вектора признаков с использованием расстояния и скорости изменения расстояния между парами движущихся точечных объектов. Эти векторы признаков затем группируются для получения компонентных атомарных событий, которые используются для описания человеческой производственной деятельности с помощью эгоцентрической камеры. Этот подход основан на известном фиксированном наборе объектов, где распознаются взаимодействия между идентификаторами маркеров запястья и набором помеченных объектов. Каждый из этих подходов основан на анализе наблюдаемых данных в автономном процессе, где отношения изучаются путем взятия всего набора данных, взаимодействий между сущностями и изучения подходящих дискретизаций для наилучшего представления данных. Цель этой работы состоит в том, чтобы показать, что постепенное обучение на протяжении всей жизни робота может существовать, поэтому качественные представления определяются вручную заранее [19]. Объектно-ориентированный и качественный процессы преобразования наблюдаемых видеоданных частично снимают проблемы, связанные с особенностями низкокачественного изображения, которое использовалось с вероятностными подходами к обучению в предыдущих работах. Другие работы непосредственно сравнивали характеристики STIP с качественными характеристиками на трех сложных эгоцентрических камерах и продемонстрировали, что качественные представления могут превосходить традиционные характеристики изображения, когда доступны положения объектов [20]. То есть качественное представление может поддерживать семантически значимые реляционные последовательности и информацию, конкретно относящуюся к движениям. Качественные характеристики, возможно, кодируют более «долгосрочную временную информацию». Это связано с общим представлением набора слов (где порядок слов часто теряется с использованием функций изображения), временное перекрытие внутри наших функций поддерживает важную структуру в наблюдении, предлагая полное преимущество разреженных, дискретных представлений. Это позволяет данному подходу изучать латентные шаблоны обычно наблюдаемых качественных признаков. Неконтролируемый подход в сочетании с единым качественным пространственным исчислением ранее использовался для кодирования непрерывных видеопоследовательностей разворота самолета. Здесь грануляр-

ность классов деятельности, изучаемых с использованием неконтролируемой техники, ограничена проблемами восприятия, связанными с абстрактными входными сценами. Тут используется одна локация камеры, но наблюдаются медленно движущиеся объекты, которым недостает изменчивости динамических движений человека. Эгоцентрическая камера используется для изучения аналогичного качественного представления движений позы человеческого тела. Однако оба эти подхода изучают деятельность в автономном и пакетном процессах, где целью работы является использование приближенных вариационных методов для решения практических вопросов, связанных с мобильной робототехникой, выполняющей пожизненное обучение человеческой деятельности.

Цель исследования также состоит в том, чтобы изучить человеческую деятельность, за которой наблюдает мобильный робот на протяжении всего времени нахождения среди людей. Сейчас опишем количественные входные данные, которые получает мобильный робот. Сначала определим, что вообще рассматривается под человеческой деятельностью, а также конкретные области деятельности, в которых должен работать робот. Затем будут представлены детали того, как робот кодирует каждое человеческое движение в виде количественной последовательности поз человеческого тела. Наконец, узнаем, как робот интерпретирует свою среду и узнает местоположения объектов, которые помогают человеку в работе.

#### *Виды человеческой деятельности*

Введем термин «активность», чтобы связать динамическую конфигурацию с несколькими «агентами», где последние могут существовать в реальности или же могут быть онлайн-агентами. В этом исследовании главное:

1) изучение человеческой деятельности как паттернов, выполняемых людьми в реальности;

2) масштабирование системы, для обеспечения непрерывного обучения мобильного робота. В этом случае будут рассматриваться только отдельные виды человеческой деятельности. Для этого вначале нужно выяснить взаимодействие между человеком и окружающей средой, а именно, между человеком и некоторыми объектами, которые обеспечивают его деятельность.

Одной из главных проблем является разрешение человеческой деятельности, которая может быть в некоторой степени ограничена имеющимся восприятием или сенсорами робота. Эта проблема является

основной для мобильного робота, поэтому его восприятие человеческой деятельности может быть ограничено его собственными сенсорами и возможностями поля зрения. Также это является самой главной проблемой, потому что современное восприятие окружающей среды роботами все еще далеко от восприятия человека. Это влияет на способность машины замечать объекты (статические или динамические) в среде, где они находятся, и изучать только паттерны активности на определенном уровне детализации. В недавних работах по изучению плана деятельности использовались обнаруженные роботом движения рук и точки их соприкосновения с объектами окружающей среды [21], для обучения робота человеческой деятельности по многочисленным видео из Интернета [22]. Однако они опираются на гораздо более узкую точку зрения, чем рассматриваемый автономный мобильный робот, в них обычно используют предварительно обученные ручные или нейронные сети для классификации.

#### *Оценка положений человека*

Мобильный робот замечает людей и сразу может сделать вывод об их 3D-позе (15 положений суставов тела), когда они появляются в поле зрения его датчика RGBD. Общепринятым подходом является использование трекера OpenNI (Open Natural Interaction) для обнаружения нескольких людей и определения их 3D-позы в онлайн-режиме. Особенно важно получить достоверные данные позы в случаях взаимодействия человека и объектов из труднодоступных мест расположения робота. К сожалению, это вызывает множество ошибок распознавания позы из OpenNI, где объект случайно может быть рассмотрен как часть человека. Чтобы решить эту проблему, можно использовать данные о цвете RGB, чтобы помочь отличать объект от человека. В этом случае система распознавания позы человека работает в двухфазном подходе. В первой фазе эффективность OpenNI используется для производства ограничительных коробок человека на кадр. Во второй фазе ограничительные рамки человека и рамка RGB подаются в качестве входных данных в современную сверточную сеть (ConvNet) двухмерного (2D) распознавания позы человека [23]. В итоге координаты положения тела OpenNI заменяются двухмерными координатами тела, предоставленными ConvNet. Используя оцененное местоположение автономного мобильного робота, представим распознавание позы человека в виде сообщения, в котором обнаруженное роботом местоположение части

тела человека представлено трехмерными декартовыми координатами в представленной роботом системе отчета вместе с соответствующим положением человека, преобразованным в глобальную систему отсчета. Тогда человеческое местоположение включает в себя набор положений суставов тела, то есть используется реализация OpenNI или ConvNet.

Для каждого человека, обнаруженного роботом, получается определенная последовательность человеческих поз, полученных в течение определенного времени работы камеры робота. Определяется последовательность человеческих поз, где каждая является обнаруженной в определенный момент времени позой, и никакие ограничения не могут быть наложены на длину записанных последовательностей. Это изменение длины является основной трудностью при использовании существующих данных для изучения деятельности человека на мобильном роботе.

#### *Представление роботом объектов окружающей среды*

Вторым важным компонентом в среде, где находится мобильный робот, являются объекты, с которыми взаимодействует человек. Здесь сфокусируемся на объектах, с которыми люди взаимодействуют в повседневной жизни и которые обеспечивают определенную работу в сфере человеческой деятельности. Например, человек может подойти к стоящему на столе компьютеру, остановиться перед ним, чтобы выполнить какое-то определенное действие (возможно, ввести на клавиатуре пароль, для того чтобы войти в систему), и при этом его части тела пространственно взаимодействуют с объектом. По этой причине представление о человеческой деятельности может включать в себя положения людей относительно объектов, с которыми они взаимодействуют, в среде работы мобильного робота. Однако обнаружение и отслеживание произвольных объектов в реальном времени с роботизированной платформы может быть сложной и не решаемой задачей. Поэтому, чтобы узнать положение интересующих человека объектов в окружающей среде, робот сначала предварительно строит 3D-модель помещения или пространства, где он сейчас находится (улица, квартира, лаборатория и т.д.), путем слияния нескольких изображений RGBD. При этом робот перемещает свой угол наклона несколько раз, тем самым захватывая изображение RGB и соответствующее облако точек глубины для каждого положения и обозначая каждый пиксель в облаке точек глубины со значени-

ем RGB из соответствующего изображения RGB. Вышеописанный процесс называют зачисткой. Для создания большого зарегистрированного облака точек представления всей среды робота проводится сразу несколько зачисток. Как только робот получает трехмерное представление облака точек местоположения его нахождения, он извлекает местоположения потенциальных объектов путем рендеринга поверхности с использованием поверхностных элементов и извлечения кластеров. Современная производительность в извлечении семантически значимых сегментов может быть достигнута с помощью алгоритма сегментации без вмешательства человека. Это показано в источнике [24], и был описан метод, аналогичный тому, который представлен в источнике [25]. Эти сегменты были рассмотрены как потенциально важные объекты в среде работы мобильного робота.

#### Результаты исследования и их обсуждение

Таким образом, была освещена новая структура, в которой изучаются мало-размерные представления наблюдений за человеческой деятельностью с помощью мобильного робота. Было показано, что, преобразуя наблюдения с использованием качественных пространственных отношений между отслеживаемыми объектами в визуальной сцене и выполняя вероятностные методы неконтролируемого обучения, можно узнать эффективные тематические распределения, представляющие человеческую деятельность. В качестве главного аспекта представлено формальное представление наблюдений за человеком, полученных мобильным роботом, качественные преобразования для их обобщения и методы извлечения дискретных признаков в виде последовательностей наблюдаемых качественных отношений. Наконец, было показано, что эта структура может хорошо работать, учитывая реальные существующие практические проблемы мобильной робототехники, о которых в настоящее время очень мало говорится. Видно, что из многочисленных наблюдений за человеком в реальных условиях мобильный робот может узнать последовательные и значимые шаблоны подробных 3D-последовательностей поз человеческого тела, используя неконтролируемые методы обучения, применяемые к новому качественному представлению человеческих наблюдений. Модели человеческой деятельности изучаются при наличии динамических объектов в наборе данных поэтапной установки статической камеры, а также в более сложной, реальной

среде с автоматическим изучением местоположения объектов с помощью трекеров OpenNI. Была рассмотрена осуществляющаяся двухфазным подходом система распознавания положения частей тела человека, а также некоторые виды деятельности человека и проблемы в распознавании мобильным роботом важнейших объектов, с которыми взаимодействует человек. Наконец, были предложены одни из возможных решений интересных и пока еще не решенных практических задач в области анализа деятельности человека и его взаимодействия с объектами окружающей среды с помощью мобильного робота, работающего в реальных условиях. Выяснено, что с помощью более сложных методов обучения можно устранить некоторые практические ограничения, связанные с обучением мобильного робота человеческой деятельности на протяжении всей жизни. А именно то, что ручная временная сегментация не требуется и что вариационный метод Байеса может быть применен для инкрементных и пожизненных настроек обучения. Одним из возможных направлений исследований в будущем могло бы стать распространение этих данных на многомесячные данные наблюдений. Это позволило бы обнаружить совершенно новые темы, возможно, от робота, входящего в совершенно новые среды, возможно, и за пределами Земли. Кроме того, параметр «скорость обучения» мог бы обновляться в режиме онлайн, учитывая новые среды, исследуемые роботом, чтобы быстрее сходиться на новых наблюдаемых действиях человека.

#### Список литературы

1. E. Marder-Eppstein, Berger E., Foote T., Gerkey B., Konolige K. The office marathon. IEEE Conference on Robotics and Automation. 2010. P. 1–8.
2. Hawes N., Duckworth P., Burbridge C., Jovan F., Kunze L., Lacerda B., Mudrova L., Young J., Wyatt J., Hebesberger D., Kortner T., Ambrus R., Bore N., Folkesson J., Jensfelt P., Beyer L., Hermans A., Leibe B., Aldoma A., Faulhammer T., Zillich M., Vincze M., Chinellato E., Al-Omari M., Gatsoulis Y., Hogg D.C., Cohn A.G., Dondrup C., Fentanes J.P., Krajnik T. The strands project: long-term autonomy in everyday environments. IEEE Robotic & Automation Magazine. 2017. vol. 24. no. 3. P. 146–156.
3. Duckworth P., Alomari M., Bore N., Hawasly M., Hogg D.C., Cohn A.G. Grounding of human environments and activities for autonomous robot. Proceedings of 26th International Joint Conference on Artificial Intelligence. 2017. P. 1395–1402.
4. Duckworth P., Alomari M., Charles J., Hogg D.C., Cohn A.G. Latent Dirichlet allocation for unsupervised activity analysis on an autonomous mobile robot. Proceedings of Thirty-first Association for the Advancement of Artificial Intelligence. 2017. P. 3819–3826.
5. Duckworth P., Hogg D.C., Cohn A.G. Unsupervised Human Activity Analysis for Intelligent Mobile Robots. Artificial Intelligence. 2019. no. 270. P. 67–92.

6. Chen L., Hoey J., Nugent C.D., Cook D.J., Yu Z. Sensor-based activity recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C Applications and Reviews*. 2012. vol. 42. no. 6. P. 790–808.
7. Lara O.D., Labrador M.A. A survey on human activity recognition using wearable sensors. *IEEE Communication Surveys & Tutorials*. 2012. vol. 15. no. 3. P. 1192–1209.
8. Weinland D., Ronfard R., Boyer E. A survey of vision-based methods for action representation, segmentation and recognition. *Computer Vision and Image Understanding*. 2011. vol. 115. no. 2. P. 224–241.
9. Ye M., Zhang Q., Wang L., Zhu J., Yang R., Gall J. A survey on human motion analysis from depth data. *Time-of-Flight and Depth Imaging: Sensors, Algorithms, and Applications*. 2013. P. 149–187.
10. Aggarwal J., Xia L. Human activity recognition from 3D data: a review. *Pattern Recognition Letters*. 2014. vol. 48. P. 70–80.
11. Duckworth P., Gatsoulis Y., Jovan F., Hawes N., Hogg D.C., Cohn A.G. Unsupervised learning of qualitative motion behaviours by a mobile robot. *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. 2016. P. 1043–1051.
12. Duckworth P., Alomari M., Gatsoulis Y., Hogg D.C., Cohn A.G. Unsupervised activity recognition using latent semantic analysis on a mobile robot. *Proceedings of 22nd European Conference on Artificial Intelligence*. 2016. P. 1062–1070.
13. Gatsoulis Y., Alomari M., Burbridge C., Dondrup C., Duckworth P., Lightbody P., Hanheide M., Hawes N., Hogg D.C., Cohn A.G. QSRLib: a software library for online acquisition of qualitative spatial relations from video. *Workshop on Qualitative Reasoning*, at IJCAI. 2016. P. 1–6.
14. Zhang J., Gong S. Action categorization by structural probabilistic latent semantic analysis. *Computer Vision and Image Understanding*. 2010. vol. 114. no. 8. P. 857–864.
15. Wu C., Zhang J., Savarese S., Saxena A. Watch-n-patch: unsupervised understanding of actions and relations. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015. P. 4362–4370.
16. Chen J., Cohn A., Liu D., Wang S., Ouyang J., Yu Q. A survey of qualitative spatial representations. *Knowledge Engineering Review*. 2015. vol. 30, no. 1. P. 106–136.
17. Tayyub J., Tavanai A., Gatsoulis Y., Cohn A.G., Hogg D.C. Qualitative and quantitative spatio-temporal relations in daily living activity recognition. *Proceedings of the 12th Asian Conference on Computer Vision*. 2015. P. 115–130.
18. Crouse M., Forbus K.D. Elementary school science as a cognitive system domain: how much qualitative reasoning is required. *Proceedings of Fourth Annual Conference on Advances in Cognitive Systems*. 2016. P. 1–12.
19. Alomari M., Duckworth P., Hogg D.C., Cohn A.G. Semi-supervised natural language acquisition and grounding for robotic systems. *Proceedings of Thirty-first Association for the Advancement of Artificial Intelligence*. 2017. P. 4349–4356.
20. Bleser G., Damen D., Behera A., Hendeby G., Mura K., Miezal M., Gee A., Petersen N., Macaeas G., Domingues H., Gorcecky D., Almeida L., W. Mayol-Cuevas, Calway A., Cohn A.G., Hogg D.C., Stricker D. Cognitive learning, monitoring and assistance of industrial workflows using egocentric sensor networks. *PLoS ONE*. 2015. vol. 10. no. 6. Article e0127769. P. 1–41. [Electronic resource]. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0127769> (date of access: 25.12.2019).
21. Fermüller C., Wang F., Yang Y., Zampogiannis K., Zhang Y., Barranco F., Pfeiffer M. Prediction of manipulation actions. *International Journal of Computer Vision*. 2018. no. 126. P. 358–374.
22. Yang Y., Li Y., Fermüller C., Aloimonos Y. Robot learning manipulation action plans by «watching» unconstrained videos from the world wide web. *Proceedings of 29th Conference AAAI on Artificial Intelligence*. 2015. P. 3686–3693.
23. Wei S., Ramakrishna V., Kanade T., Sheikh Y. Convolutional pose Machines. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. 2016. P. 4724–4732.
24. Schoeler M., Papon J., Worgotter F. Constrained planar cuts-object partitioning for point clouds. *The IEEE Conference on Computer Vision and Pattern Recognition*. 2015. P. 5207–5215.
25. Bore N., Ambrus R., Jensfelt P., Folkesson J. Efficient retrieval of arbitrary objects from long-term robot observations. *Robotic and Autonomous Systems*. 2017. no. 91. P. 139–150.