

doi: 10.17586/2226-1494-2021-21-4-571-577

УДК 004.942

## Исследование движения человека в системах компьютерного зрения на основе скелетной модели

Софья Алексеевна Казакова<sup>1</sup>, Полина Андреевна Леонтьева<sup>2</sup>,  
Мария Ильинична Фролова<sup>3</sup>, Юлия Валерьевна Донецкая<sup>4</sup>,  
Илья Юрьевич Попов<sup>5</sup>✉, Александр Юрьевич Кузнецов<sup>6</sup>

<sup>1,2,3,4,5,6</sup> Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

<sup>1</sup> [sophie.a.kaz@gmail.com](mailto:sophie.a.kaz@gmail.com), <https://orcid.org/0000-0001-7264-8189>

<sup>2</sup> [polinleonteva@gmail.com](mailto:polinleonteva@gmail.com), <https://orcid.org/0000-0002-9423-0266>

<sup>3</sup> [talviveden@gmail.com](mailto:talviveden@gmail.com), <https://orcid.org/0000-0002-6043-7445>

<sup>4</sup> [donetskaya\\_julia@mail.ru](mailto:donetskaya_julia@mail.ru), <https://orcid.org/0000-0001-5293-5025>

<sup>5</sup> [ilyapopov27@gmail.com](mailto:ilyapopov27@gmail.com)✉, <https://orcid.org/0000-0002-6407-7934>

<sup>6</sup> [al.ur.kouznetsov@gmail.com](mailto:al.ur.kouznetsov@gmail.com), <https://orcid.org/0000-0002-5702-3786>

### Аннотация

**Предмет исследования.** Анализ движения человека в системах компьютерного зрения выполняется как в двухмерном, так и в трехмерном пространствах. Особенность анализа фигур в двухмерном пространстве состоит в применении изображения с одной камеры и/или множественных нателных датчиков. Подобный подход приводит к быстрому накоплению ошибки и, как следствие, низкой точности представления фигуры. В случае выполнения трехмерного анализа используются несколько камер, а фигуры представлены в виде совокупностей объемных элементов. Несмотря на высокую точность трехмерного анализа, он связан с большой вычислительной сложностью и нагрузкой на внутреннюю компьютерную сеть. Задача исследования состоит в разработке модели, с использованием изображения одной камеры. При этом поставлена задача достижения точности, приближенной к трехмерному анализу. **Метод.** В рассмотренной модели фигура человека представляется в виде скелета, описанного ациклическим связным графом. Выполнен анализ структуры человеческой фигуры и выделены пятнадцать основных точек. Исследовано и дано математическое описание физических и логических связей. Скоростные и пространственные характеристики точек и связей описывают общую динамику движения. **Основные результаты.** Представлено описание модели движения человека и приведен пример ее построения для конкретного изображения. На основе предложенной модели разработан алгоритм сбора и анализа информации о взаимных расположениях и скоростных характеристиках узлов и ребер графа. **Практическая значимость.** Разработанная модель может быть использована для сбора информации об эталонной (нормальной) динамике движения человека. В случае детектирования динамики, отклоняющейся от эталонной, поведение определяется как девиантное. Предложенный алгоритм может найти применение в системах компьютерного зрения для детектирования и анализа движения человека, а также выявления отклонений движений от эталонных.

### Ключевые слова

компьютерное зрение, анализ движения человека, поведенческая аналитика, детектирование движения, скелетная модель

**Ссылка для цитирования:** Казакова С.А., Леонтьева П.А., Фролова М.И., Донецкая Ю.В., Попов И.Ю., Кузнецов А.Ю. Исследование движения человека в системах компьютерного зрения на основе скелетной модели // Научно-технический вестник информационных технологий, механики и оптики. 2021. Т. 21, № 4. С. 571–577. doi: 10.17586/2226-1494-2021-21-4-571-577

## A study of human motion in computer vision systems based on a skeletal model

Sophia A. Kazakova<sup>1</sup>, Polina A. Leonteva<sup>2</sup>, Maria I. Frolova<sup>3</sup>,  
Julia V. Donetskaya<sup>4</sup>, Ilya Yu. Popov<sup>5</sup>, Aleksandr Yu. Kuznetsov<sup>6</sup>

1, 2, 3, 4, 5, 6 ITMO University, Saint Petersburg, 197101, Russian Federation

<sup>1</sup> sophie.a.kaz@gmail.com, <https://orcid.org/0000-0001-7264-8189>

<sup>2</sup> polinleonteva@gmail.com, <https://orcid.org/0000-0002-9423-0266>

<sup>3</sup> talviveden@gmail.com, <https://orcid.org/0000-0002-6043-7445>

<sup>4</sup> donetskaya\_julia@mail.ru, <https://orcid.org/0000-0001-5293-5025>

<sup>5</sup> ilyapopov27@gmail.com, <https://orcid.org/0000-0002-6407-7934>

<sup>6</sup> al.ur.kouznetsov@gmail.com, <https://orcid.org/0000-0002-5702-3786>

### Abstract

Methods of studying human motion in computer vision systems can be divided into two types. These are analysis in two-dimensional and three-dimensional space. The former uses a single camera image and/ or multiple body sensors. Such an approach leads to a rapid accumulation of error and, consequently, low accuracy of the figure representation. Multiple cameras are usually used in the case of three-dimensional space analysis, while the objects are represented as sets of volumetric elements. Despite the high accuracy of this method, it is associated with high computational complexity and internal network load. The purpose of the paper is to develop a model using a single camera, while approaching three-dimensional space analysis methods in terms of accuracy. In this paper a human figure is represented as a skeleton. The skeleton is described by an acyclic connected graph. The general structure of a human figure is analyzed. Fifteen basic points are selected. Physical and logical connections between them were studied and mathematically described. The velocity and spatial characteristics of the points and connections outline the general dynamics of motion. The study describes a model of human motion and gives the option for model construction on the example of a particular image. The developed algorithm for collection and analysis of information estimates relative locations and velocity characteristics of the graph elements. The model can be used for acquisition of information about the reference dynamics of human movements. In case of detecting major differences between the reference and the reality, the behavior is defined as deviant. Thus, the obtained algorithm can be applied in computer vision systems for detection and analysis of human movements.

### Keywords

computer vision, human motion analysis, behavioral analytics, motion detection, skeletal model

**For citation:** Kazakova S.A., Leonteva P.A., Frolova M.I., Donetskaya Ju.V., Popov I.Yu., Kuznetsov A.Yu. A study of human motion in computer vision systems based on a skeletal model. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2021, vol. 21, no. 4, pp. 571–577 (in Russian). doi: 10.17586/2226-1494-2021-21-4-571-577

### Введение

Движение человека можно представить в виде набора траекторий и последовательных изменений положения корпуса и конечностей человека в пространстве, которые связаны с определенными факторами окружающей среды [1]. Факторы можно разделить на внешние и внутренние. К внешним относятся одежда, наклон поверхности, природные явления. Внутренние – мускулатура, скелет, травмы, вес, привычки человека и др. Перечисленные факторы позволяют человеку определенным образом реагировать на изменения в среде, создавая при этом поведенческую картину действий для перемещения частей тела. Благодаря схожести реакций людей на те или иные события, становится возможным анализ движения человека для дальнейшей обработки информации и предсказания его поведения.

Видеоаналитика движения человека за последние годы стала неотъемлемой частью исследований в сфере компьютерного зрения, и является на данный момент передовой областью развития подобных систем. Интерес к средствам компьютерного зрения и анализу движения человека, в частности, мотивирован широким спектром открывающихся возможностей: организация систем безопасности, анализ спортивных результатов, человеко-машинные интерфейсы, видеонаблюдение и видеоконференцсвязь.

Алгоритмы распознавания движений с помощью систем компьютерного зрения подобно человеческо-

му глазу должны фиксировать изменения координат в двухмерном или трехмерном пространствах и зависят от метода построения математической модели. Выбор формата рассматриваемого изображения — первое и основное различие предложенных исследователями ранних подходов.

Существующие подходы к исследованию и регистрации изменения положения человека в пространстве можно разделить по ряду признаков:

- тип используемых моделей [2] (на основе скелетной формы [3, 4], объемных фигур [5, 6] и др.);
- модальность датчиков [7];
- множественность датчиков и камер (монокулярный [8] в отличие от стерео [9]);
- размещение датчиков (централизованное — распределенного);
- мобильность датчиков (стационарное — движущегося [10]);
- основная (практически применяемая) модель [11].

Одна из основных проблем анализа движения человека – применение сложных алгоритмов отслеживания в распознавании действий. Для решения данной проблемы исследователям необходимо использовать большие вычислительные мощности для разработки систем видеоаналитики [12]. Следовательно, крайне актуальной является задача разработки математической модели, которая могла бы нивелировать недостатки существующих методов путем нахождения оптимального совмещения разных моделей анализа человеческой

фигуры и при этом снизить вычислительную нагрузку систем компьютерного зрения.

Рассмотренные математические модели имеют ряд недостатков. В решениях, использующих алгоритмы, основанные на двухмерном изображении и отслеживании точек фигуры, существуют проблемы, связанные с накоплением ошибки, вытекающей из четкого представления кинематической цепи (совокупности точек, связанных ребрами). Так, в работе, представленной Манон Кок, Карстеном Экхоффом, Ивом Вейгерсом, Томасом Силом [13], демонстрируется скелетная модель на примере эксперимента с гироскопическими датчиками, закрепленными на ноге. Анализ ведется по двум жестко связанным узлам скелетной модели. Такой подход приводит к накоплению ошибок и погрешности в несколько градусов, что ведет к низкой точности работы алгоритма, непосредственно влияющей на производительность и результативность процесса отслеживания. Это происходит из-за неучета вариативности положения датчиков (или, в иных случаях, камер) при детектировании и анализе движения человека.

В классических методах многообъектного слежения используются жесткие алгоритмы кластерного анализа (например, алгоритм  $K$ -средних), сопряженные с высокой вычислительной сложностью и, как следствие, значительными затратами ресурсов. В работе Даниэла Эрикссона и Йонаса Харстрема [14] представлена новейшая разработка детектирования движения с помощью системы LiDAR. На данный момент это один из самых оптимальных вариантов анализа движения, однако он не применяется для детектирования людей, поскольку затраты на аппаратное обеспечение не окупаются при подобном использовании. Более того, подобные алгоритмы часто имеют ошибки в случае появления в кадре сложных преград, ограничивающих обзор камеры и перекрывающих части отслеживаемых фигур.

Вопрос эффективного построения скелетной модели, не раз рассмотренный российскими исследователями, предлагает вариативную классификацию поз на фотографии с помощью методов машинного обучения. Такая классификация не служит полноценным решением распознавания аномальных движений или явлений. Она является первой ступенью в решении данного вопроса, которая не подразумевает полноценного анализа математических и статистических параметров положения узловых точек скелетной модели [15, 16].

Математическая модель, рассмотренная в настоящей работе, служит основой для разработки и построения алгоритма видеоаналитики движения человека, реализующегося средствами компьютерного зрения. Алгоритм должен определять тип движения фигуры и отслеживать девиации — отклонения поведения наблюдаемой фигуры от эталонного, заданного исходной математической моделью.

Цель разрабатываемой модели — создание высокопроизводительной системы безопасности, основным элементом которой является модуль видеонаблюдения с применением компьютерного зрения, позволяющий определять различные виды девиации поведения человека. Реализация планируется в псевдореальном мас-

штабе времени с применением алгоритмов нейронной сети.

В задачи модели входит детектирование человека в кадре, построение скелетной модели по фигуре и отслеживание ее движений. При этом скоростные характеристики перемещения и изменения взаимного расположения точек модели определяют тип происходящего отклонения. Такое решение не основывается на определенной размерности пространства слежения, что является неоспоримым преимуществом перед существующими моделями, повышая точность результатов.

### Построение скелетной модели человека

Объектом моделирования в настоящем исследовании является движение человека. Цель моделирования — выявление аномалий в видеопотоке для детектирования инцидента в поведении наблюдаемого субъекта и дальнейшей передачи информации о нем оператору системы видеонаблюдения.

Рассмотрим первый этап разработки модели, упрощающий фигуру человека до связного ациклического графа — дерева с корнем в верхней точке фигуры (затылок) и листьями (в конечностях). Промежуточными вершинами являются колени, три точки таза (центр и тазобедренные суставы), шея, плечевые и локтевые суставы.

*Опорными* в настоящей работе назовем точки-узлы дерева.

Математическая модель движения человека представляется в виде комбинации из четырех компонентов — математических моделей отдельных структур: опорных точек; смежности точек; расположения; движения.

Модель опорных точек представляет собой вектор-столбец  $\mathbf{P}$  с координатами опорных точек  $P_1, P_2, \dots, P_{15}$ :

$$\mathbf{P} = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_{15} \end{bmatrix} = \begin{bmatrix} (x_1, y_1) \\ (x_2, y_2) \\ \vdots \\ (x_{15}, y_{15}) \end{bmatrix}, \quad (1)$$

где  $x_1, x_2, \dots, x_{15}$  — координаты опорных точек по оси  $OX$ ;  $y_1, y_2, \dots, y_{15}$  — по оси  $OY$ .

$XOY$  — система координат, в которой производится построение двумерной проекции модели. Для каждой фигуры в кадре строится собственная система координат  $XOY_\omega$ , плоскости построения при этом параллельны друг другу, базисы выбираются на основе анализа глубины изображения.

Положения  $P_1, P_2, \dots, P_{15}$  соответствуют схеме на рис. 1, а.

Модель смежности точек описывает ребра графа и представляет собой квадратную матрицу  $\mathbf{X}(\mathbf{P})_{15 \times 15}$ , элементами которой являются значения функции вида:

$$\Delta_{ij} = \begin{cases} 0, & \text{если } P_i \text{ и } P_j \text{ не соединяются ребром,} \\ 1, & \text{в обратном случае.} \end{cases}$$

Тогда  $\mathbf{X}(\mathbf{P}) = [\Delta_{ij}]$ , т. е.

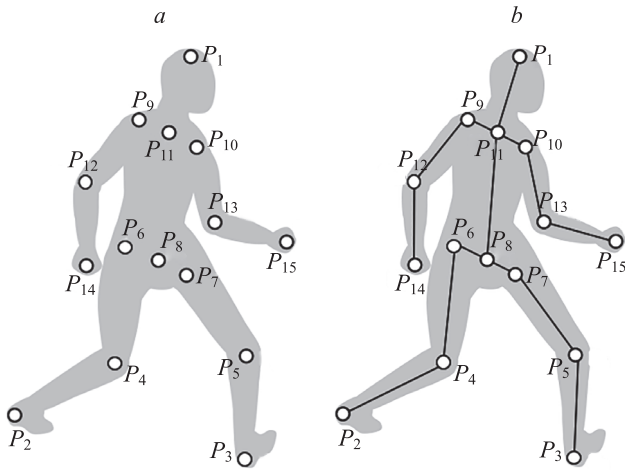


Рис. 1. Расположение точек на фигуре человека (a); узлы дерева и связывающие их ребра (b)

Fig. 1. Scheme of the graph nodes on a human figure (a); the tree nodes and edges connecting them (b)

$$\mathbf{X}(\mathbf{P}) = \begin{bmatrix} \Delta_{1,1} & \Delta_{1,2} & \dots & \Delta_{1,15} \\ \Delta_{2,1} & \Delta_{2,2} & \dots & \Delta_{2,15} \\ \dots & \dots & \ddots & \dots \\ \Delta_{15,1} & \Delta_{15,2} & \dots & \Delta_{15,15} \end{bmatrix} \quad (2)$$

Описываемый данной моделью идеальный граф представлен на рис. 1, b. Тогда матрица смежности  $\mathbf{X}(\mathbf{P})$  имеет вид

$$\mathbf{X}(\mathbf{P}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \quad (3)$$

Матрица описывает эталонный случай детектирования опорных точек, когда все 15 из них находятся в поле зрения камеры. Тот же эталонный случай графически изображен на рис. 1, b.

Так, например, в первой строке единицы стоят в первом и одиннадцатом столбце, это означает, что первая точка  $P_1$  связана только с  $P_{11}$ ; с  $P_1$  она, очевидно, совпадает. Аналогичным образом можно рассмотреть все остальные строки матрицы.

Модель расположения описывает изменение положения опорных точек в пространстве и имеет вид

матрицы  $\mathbf{U}(\mathbf{P}) = [u_{ij}(P_i)]$ , где  $u_{ij}(P_i)$  — координаты точек  $P_i$ , считываемые с каждого кадра видеопотока во времени  $t$ .

Тогда :

$$\mathbf{U}(\mathbf{P}) = \begin{bmatrix} u_{1,1}(P_1) & u_{1,2}(u_{1,1}(P_1)) & \dots & u_{1,t}(u_{1,t-1}(P_1)) \\ u_{2,1}(P_2) & u_{2,2}(u_{2,1}(P_2)) & \dots & u_{2,t}(u_{2,t-1}(P_2)) \\ \dots & \dots & \ddots & \dots \\ u_{15,1}(P_{15}) & u_{15,2}(u_{15,1}(P_{15})) & \dots & u_{15,t}(u_{15,t-1}(P_{15})) \end{bmatrix} = \begin{bmatrix} U_1(P_1) \\ U_2(P_2) \\ \vdots \\ U_{15}(P_{15}) \end{bmatrix} \quad (4)$$

Модель движения описывает непосредственно поведение фигуры и представляет собой вектор-столбец  $\mathbf{F}(\mathbf{P}) = [F_i(P_i)]$ , где  $F_i(P_i)$  — функция, описывающая движение каждой из точек  $P_1, P_2, \dots, P_{15}$ . Определение системой идентификации точки  $P_i$  будет рассмотрено в следующих работах.

Порядок вывода функции движения точки следующий.

1. Изменение положения точки  $P_i$  фиксируется в массиве  $U_i(P_i)$ . Массивы  $U_1(P_1), U_2(P_2), \dots, U_{15}(P_{15})$  — списки последовательных значений координат во времени. Количество записанных значений при этом зависит от динамики движения точки и от длительности ее нахождения в кадре: чем чаще меняются координаты, тем больше значений фиксируется. Дублирующиеся значения затираются по причине избыточности данной информации для дальнейшей аналитики.
2. Функция  $F_i(P_i)$  заданная набором точек  $U_i(P_i)$ , аппроксимируется гармоническим рядом для получения гармонической функции, описывающей движения данной точки. Результаты аппроксимации записываются в вектор-столбец  $F(\mathbf{P})$ . Допустим, что представленная в табличном виде функция  $F_i(P_i)$  кусочно-непрерывна, кусочно-монотонна и ограничена, тогда  $F_i(P_i)$  удовлетворяет условиям Дирихле, и, следовательно, может быть выполнена аппроксимация сплайн-функцией, или, как представлено в (3), аппроксимация нечетными функциями.
3. Быстрое изменение координаты точки во времени, происходящее при резких движениях отслеживаемой фигуры, влечет за собой значительный скачок амплитуды синусоиды. Для отслеживания подобных выбросов, которые являются первичными признаками девиантного поведения в разрабатываемой математической модели, на каждом шаге аппроксимации проверяется производная в последней зафиксированной точке функции  $F_i'(U_{i-1})$ .

Рассмотрим реализацию построения связного ациклического графа на примере конкретного изображения.

### Пример реализации

На этапе предобработки видеопотока движущееся тело выделяется по контуру на основе покадрового попиксельного сравнения изображений методом динамической сегментации [17] (рис. 2, *a*). Дальнейшее построение графа основывается на анализе границ контура выделенной области и вычислении градиента по плотности распределения пикселей внутри этой области.

В первую очередь строится корень дерева. Он определяется как верхняя точка выделенного контура по оси  $Y$ . Затем выделяются нижние опорные точки контура, которые определяются как нижние по  $Y$  — листья дерева. Между ними по оси  $X$  должна находиться точка излома контура, которая выделяется как промежуточная опорная точка (рис. 2, *b*).

По градиенту от нижних точек идет построение следующего узла дерева. В зависимости от взаимного положения с точкой излома устанавливается, является ли точка тазобедренным или коленным суставом. Если излом находится ниже найденной точки, то она определяется тазобедренным суставом или коленом. Между точками тазобедренных суставов находится опорная точка крестца. Ее положение формируется проекцией точки излома контура на ребро таза. Если найдены тазобедренные суставы, то точки ступней и бедер соединяются ребрами. По нескольким кадрам движения находятся точки коленей как изменения направления вектора, разбивающие ребра ног на ребра голени и бедер. В случае, если найдены точки коленей, тазобедренные суставы находятся также по градиенту, дальнейшие преобразования аналогичны.

Положение точек тазобедренных суставов позволяет найти точки плечевых суставов, расположенные в верхней части контура, по перпендикуляру вверх до границ контура. В дальнейшем движения фигуры

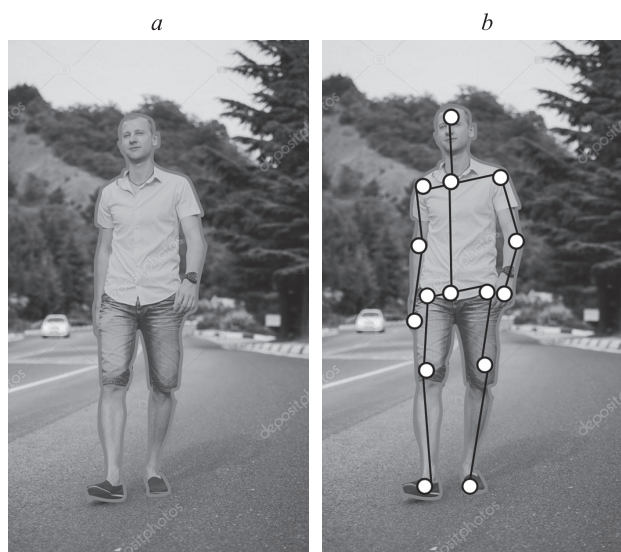


Рис. 2. Изображения с выделенной областью фигуры (*a*) и после построения математической модели (*b*)

Fig. 2. Image with a highlighted figure (*a*); image with a constructed graph (*b*)

дают возможность скорректировать положения этих точек. Алгоритм строит «черновую версию» по первому кадру на основе анализа статичного изображения, и далее уточняет положения суставов при динамическом анализе. Точки плечевых суставов соединяются ребром плечевого пояса. Под точкой головы на ребре плечевого пояса находится опорная точка шеи. Точка шеи и точка крестца образуют ребро позвоночника. По контуру изображения идентифицируются опорные точки кистей рук. Построение точек локтевых суставов аналогично поиску коленных суставов, и выполняется методом градиентного анализа.

В ходе дальнейшей работы планируется реализовать данную модель в MatLab. Алгоритм, имплементирующий данную модель, будет реализован на языке программирования Python с использованием OpenCV и других библиотек, необходимых для построения нейронных сетей.

Обнаружение человека в кадре возможно при оценивании особого критерия — количества пикселей, приходящихся на реальную размерную единицу целевого объекта. В качестве «реальной размерной единицы» берется метр. Другими словами, каждый реальный метр целевого объекта должен быть определен минимальным набором пикселей. Для обнаружения человека необходимо обеспечить наличие 20 пикселей в кадре на каждый реальный метр целевого объекта [18].

Оценка качества модели связана с оценкой качества имплементирующего ее алгоритма. За основные показатели принимаются вероятности ошибок первого и второго рода: модель не определяется как фигура, являющаяся человеком, и может определить человека как случайный фрагмент видеопотока.

### Алгоритм анализа движения

Анализ движения производится по следующему алгоритму:

- детектирование движения в кадре;
- выделение человеческой фигуры в видеопотоке на основе попиксельного сравнения кадров;
- построение осей координат для данной фигуры по перпендикуляру от верхней точки силуэта;
- построение моделей (1), (2) и (4) по выделенной фигуре. В случае перекрытия части фигуры препятствием, модель достраивается по эталону, чтобы избежать ложноположительных срабатываний. На этом этапе формируется набор сплайн-функций. Анализ движения основывается на анализе этих постоянно меняющихся во времени функций;
- в случае, если значения отличаются от эталонных, поведение определяется как девиантное. Отклонения выражены как абнормальные скачки в функциях движения точек скелетной модели или резкие изменения углов между ребрами;
- все случаи, определенные системой как девиантное поведение, дополнительно анализируются человеком, полученная информация в дальнейшем используется для обучения алгоритма.

На начальном этапе планируется выявлять случаи агрессивного поведения и недомоганий: судороги, об-

мороки и т. п. Все подобные случаи объединены атипичной динамикой человеческого движения, которую можно выделить из общего потока.

### Заключение

В работе исследовано движение человека в системах компьютерного зрения на основе скелетной модели. Модель совмещает в себе сравнительно низкую ресурсозатратность и понижение влияния накопительных ошибок. В модели учтено, что в зависимости от расположения камеры и динамики движения человека, расстояния между проекциями опорных точек может меняться. Отказ от использования в модели жестких ребер уменьшает накопление ошибки. Подобный подход предлагает выгодный компромисс между высокими вычислительными мощностями и точностью результатов за счет возможного применения ее совместно с методами машинного обучения и возможностью нахождения точных параметров для безошибочного определения расположения фигуры в пространстве. Анализ изображения в двухмерной плоскости позволяет использовать

изображения только с одной камеры. Пример построения представлен на рис. 2.

На следующем этапе работы планируется интерпретировать полученную модель в среде MatLab для дальнейшего построения алгоритма выявления фигуры человека и анализа движения. Необходимо проанализировать численные результаты выделенных в данной работе характеристик. Итоговую модель предполагается использовать для создания алгоритма отслеживания и аналитики движения человека в видеопотоке на языке Python. Подобный алгоритм станет основой реализации нового поколения систем видеoaналитики, использующих крайвые вычисления.

В современном мире область применения анализа поведения активно расширяется. Одним из наиболее востребованных случаев является потребность выявления девиантного поведения в местах скопления людей, например, на железнодорожной платформе. Для реализации алгоритма, способного точно и быстро определять деструктивное поведение в подобной ситуации, необходима подобная модель детектирования и анализа движения отдельного человека.

### Литература

1. Valčík J. Similarity models for human motion data: Ph.D. Thesis. Brno: Masaryk University, 2016 [Электронный ресурс]. URL: <https://is.muni.cz/th/wx926/thesis.pdf>, свободный. Яз. англ. (дата обращения: 07.04.2021).
2. Rogez G., Weinzaepfel P., Schmid C. LCR-Net++: Multi-person 2D and 3D pose detection in natural images // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020. V. 42. N 5. P. 1146–1161. <https://doi.org/10.1109/TPAMI.2019.2892985>
3. Ke Q., Bennamoun M., An S., Sohel F., Boussaid F. A new representation of skeleton sequences for 3D action recognition // *Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. P. 4570–4579. <https://doi.org/10.1109/CVPR.2017.486>
4. Vox J.P., Wallhoff F. Preprocessing and normalization of 3D-skeleton-data for human motion recognition // *Proc. IEEE Life Sciences Conference (LSC)*. Montreal, QC, Canada. 2018. P. 279–282. <https://doi.org/10.1109/LSC.2018.8572153>
5. Shin S., Halilaj E. Multi-view human pose and shape estimation using learnable volumetric aggregation // *arXiv.org*. 2020. arXiv:2011.13427.
6. Innmann M., Zollhofer M., Nießner M., Theobalt C., Stamminger M. Volumedeform: Real-time volumetric non-rigid reconstruction // *Lecture Notes in Computer Science*. 2016. V. 9912. P. 362–379. [https://doi.org/10.1007/978-3-319-46484-8\\_22](https://doi.org/10.1007/978-3-319-46484-8_22)
7. Liu Y., Wang K., Li G., Lin L. Semantics-aware adaptive knowledge distillation for sensor-to-vision action recognition // *IEEE Transactions on Image Processing*. 2021. V. 30. P. 5573–5588. <https://doi.org/10.1109/TIP.2021.3086590>
8. Xiang D., Joo H., Sheikh Y. Monocular total capture: Posing face, body, and hands in the wild // *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019. P. 10957–10966. <https://doi.org/10.1109/CVPR.2019.01122>
9. Tanke J., Gall J. Iterative greedy matching for 3D human pose tracking from multiple views // *Lecture Notes in Computer Science*. 2019. V. 11824. P. 537–550. [https://doi.org/10.1007/978-3-030-33676-9\\_38](https://doi.org/10.1007/978-3-030-33676-9_38)
10. Elanattil S., Moghadam P. Synthetic data for non-rigid 3D reconstruction using a moving RGB-D camera // *CSIRO, Data Collection*. 2018. V. 2. <https://doi.org/10.25919/5b7b60176d0cd>
11. Wang Q. A Survey of visual analysis of human motion and its applications // *arXiv.org*. 2016. arXiv:1608.00700.
12. Aggarwal J., Cai Q. Human motion analysis: A Review // *Computer Vision and Image Understanding*. 1999. V. 73. N 3. P. 428–440. <https://doi.org/10.1006/cviu.1998.0744>

### References

1. Valčík J. Similarity models for human motion data. Ph.D. Thesis. Brno: Masaryk University, 2016. Available at: <https://is.muni.cz/th/wx926/thesis.pdf> (accessed: 07.04.2021).
2. Rogez G., Weinzaepfel P., Schmid C. LCR-Net++: Multi-person 2D and 3D pose detection in natural images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, vol. 42, no. 5, pp. 1146–1161. <https://doi.org/10.1109/TPAMI.2019.2892985>
3. Ke Q., Bennamoun M., An S., Sohel F., Boussaid F. A new representation of skeleton sequences for 3D action recognition. *Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4570–4579. <https://doi.org/10.1109/CVPR.2017.486>
4. Vox J.P., Wallhoff F. Preprocessing and normalization of 3D-skeleton-data for human motion recognition. *Proc. IEEE Life Sciences Conference (LSC)*. Montreal, QC, Canada. 2018, pp. 279–282. <https://doi.org/10.1109/LSC.2018.8572153>
5. Shin S., Halilaj E. Multi-view human pose and shape estimation using learnable volumetric aggregation. *arXiv.org*, 2020, arXiv:2011.13427
6. Innmann M., Zollhofer M., Nießner M., Theobalt C., Stamminger M. Volumedeform: Real-time volumetric non-rigid reconstruction. *Lecture Notes in Computer Science*, 2016, vol. 9912, pp. 362–379. [https://doi.org/10.1007/978-3-319-46484-8\\_22](https://doi.org/10.1007/978-3-319-46484-8_22)
7. Liu Y., Wang K., Li G., Lin L. Semantics-aware adaptive knowledge distillation for sensor-to-vision action recognition. *IEEE Transactions on Image Processing*, 2021, vol. 30, pp. 5573–5588. <https://doi.org/10.1109/TIP.2021.3086590>
8. Xiang D., Joo H., Sheikh Y. Monocular total capture: Posing face, body, and hands in the wild. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 10957–10966. <https://doi.org/10.1109/CVPR.2019.01122>
9. Tanke J., Gall J. Iterative greedy matching for 3D human pose tracking from multiple views. *Lecture Notes in Computer Science*, 2019, vol. 11824, pp. 537–550. [https://doi.org/10.1007/978-3-030-33676-9\\_38](https://doi.org/10.1007/978-3-030-33676-9_38)
10. Elanattil S., Moghadam P. Synthetic data for non-rigid 3D reconstruction using a moving RGB-D camera. *CSIRO, Data Collection*, 2018, vol. 2. <https://doi.org/10.25919/5b7b60176d0cd>
11. Wang Q. A survey of visual analysis of human motion and its applications. *arXiv.org*, 2016, arXiv:1608.00700.
12. Aggarwal J., Cai Q. Human motion analysis: A Review. *Computer Vision and Image Understanding*, 1999, vol. 73, no. 3, pp. 428–440. <https://doi.org/10.1006/cviu.1998.0744>

13. Kok M., Eckhoff K., Weygers I., Seel T. Observability of the relative motion from inertial data in kinematic chains // arXiv.org. 2021. arXiv:2102.02675.
14. Eriksson D., Harstrom J. Object detection by cluster analysis on 3D-points from a LiDAR sensor. Master's thesis in Systems, Control and Mechatronics. Chalmers University of Technology, Sweden, 2019 [Электронный ресурс]. URL: <https://odr.chalmers.se/bitstream/20.500.12380/257323/1/257323.pdf>, свободный. Яз. англ. (дата обращения: 07.04.2021).
15. Егоров Ю.А. Исследование эффективности применения классических подходов для решения задачи классификации поз человека с использованием скелетной модели // Информационные технологии и системы: труды Седьмой Всероссийской научной конференции с международным участием. 2019. С. 148–151.
16. Катаев М.Ю., Катаева Н.Г., Коробко А.П., Шаймарданов Т.М. Методика построения фронтальной скелетной модели фигуры человека по изображениям // Доклады Томского государственного университета систем управления и радиоэлектроники. 2017. Т. 20. № 4. С. 109–112. <https://doi.org/10.21293/1818-0442-2017-20-4-109-112>
17. Ваганов С.Е. Алгоритм динамической сегментации пары последовательных кадров // Компьютерная оптика. 2019. Т. 43. № 1. С. 83–89. <https://doi.org/10.18287/2412-6179-2019-43-1-83-89>
18. Driggers R.G., Cox P.G., Kelley M. National imagery interpretation rating system and the probabilities of detection, recognition, and identification // Optical Engineering. 1997. V. 36. N 7. P. 1952–1959. <https://doi.org/10.1117/1.601381>
13. Kok M., Eckhoff K., Weygers I., Seel T. Observability of the relative motion from inertial data in kinematic chains. *arXiv.org*, 2021, arXiv:2102.02675.
14. Eriksson D., Harstrom J. *Object detection by cluster analysis on 3D-points from a LiDAR sensor. Master's thesis in Systems, Control and Mechatronics*. Chalmers University of Technology, Sweden, 2019. Available at: <https://odr.chalmers.se/bitstream/20.500.12380/257323/1/257323.pdf> (accessed: 07.04.2021).
15. Egorov Y.A. Research of effectiveness of classical approaches for solving the problem of human pose classification using skeletal model. *Information Technologies and Systems. 8th Annual International Workshop*, 2019, pp. 148–151. (in Russian)
16. Kataev M.Yu., Kataeva N.G., Korobko A.P., Shaymardanov T.M. Methodology to build a frontal skeletal model of a human figure during walking using images. *Proceedings of TUSUR University*, 2017, vol. 20, no. 4, pp. 109–112. (in Russian). <https://doi.org/10.21293/1818-0442-2017-20-4-109-112>
17. Vaganov S. E. A method for dynamic segmentation of a pair of sequential video-frames. *Computer Optics*, 2019, vol. 43, no. 1, pp. 83–89. (in Russian). <https://doi.org/10.18287/2412-6179-2019-43-1-83-89>
18. Driggers R.G., Cox P.G., Kelley M. National imagery interpretation rating system and the probabilities of detection, recognition, and identification. *Optical Engineering*, 1997, vol. 36, no. 7, pp. 1952–1959. <https://doi.org/10.1117/1.601381>

#### Авторы

**Казакова Софья Алексеевна** — студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0001-7264-8189>, [sophie.a.kaz@gmail.com](mailto:sophie.a.kaz@gmail.com)

**Леонтьева Полина Андреевна** — студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-9423-0266>, [polinleonteva@gmail.com](mailto:polinleonteva@gmail.com)

**Фролова Мария Ильинична** — студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-6043-7445>, [mfrolovameltz@gmail.com](mailto:mfrolovameltz@gmail.com)

**Донецкая Юлия Валерьевна** — кандидат технических наук, доцент, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57200259228](https://orcid.org/0000-0001-5293-5025), <https://orcid.org/0000-0001-5293-5025>, [donetskaya\\_julia@mail.ru](mailto:donetskaya_julia@mail.ru)

**Попов Илья Юрьевич** — кандидат технических наук, ассистент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57202195632](https://orcid.org/0000-0002-6407-7934), <https://orcid.org/0000-0002-6407-7934>, [ilyapopov27@gmail.com](mailto:ilyapopov27@gmail.com)

**Кузнецов Александр Юрьевич** — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57195326635](https://orcid.org/0000-0002-5702-3786), <https://orcid.org/0000-0002-5702-3786>, [al.ur.kouznetsov@gmail.com](mailto:al.ur.kouznetsov@gmail.com)

#### Authors

**Sophia A. Kazakova** — Student, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0001-7264-8189>, [sophie.a.kaz@gmail.com](mailto:sophie.a.kaz@gmail.com)

**Polina A. Leonteva** — Student, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-9423-0266>, [polinleonteva@gmail.com](mailto:polinleonteva@gmail.com)

**Maria I. Frolova** — Student, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-6043-7445>, [mfrolovameltz@gmail.com](mailto:mfrolovameltz@gmail.com)

**Julia V. Donetskaya** — PhD, Associate Professor, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57200259228](https://orcid.org/0000-0001-5293-5025), <https://orcid.org/0000-0001-5293-5025>, [donetskaya\\_julia@mail.ru](mailto:donetskaya_julia@mail.ru)

**Ilya Yu. Popov** — PhD, Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57202195632](https://orcid.org/0000-0002-6407-7934), <https://orcid.org/0000-0002-6407-7934>, [ilyapopov27@gmail.com](mailto:ilyapopov27@gmail.com)

**Aleksandr Yu. Kuznetsov** — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57195326635](https://orcid.org/0000-0002-5702-3786), <https://orcid.org/0000-0002-5702-3786>, [al.ur.kouznetsov@gmail.com](mailto:al.ur.kouznetsov@gmail.com)

Статья поступила в редакцию 17.05.2021

Одобрена после рецензирования 06.07.2021

Принята к печати 01.08.2021

Received 17.05.2021

Approved after reviewing 06.07.2021

Accepted 01.08.2021



Работа доступна по лицензии  
Creative Commons  
«Attribution-NonCommercial»