Министерство науки и высшего образования Российской Федерации

Санкт-Петербургский политехнический университет Петра Великого

Институт машиностроения, материалов и транспорта

Высшая школа автоматизации и робототехники

**НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА**

**Разработка системы захвата движений человека на основе веб-камер**

по направлению подготовки 15.03.06 «Мехатроника и робототехника»

код и наименование направления подготовки (специальности)

направленность (профиль) 15.03.06\_04 «Автономные роботы»

код и наименование направленности (профиля) образовательной программы

Выполнил <*подпись*> М. И. Скрыль

студент гр. 3331506/90401

Руководитель <*подпись*> М. С. Ананьевский

доцент,

кандидат физико-математических

наук

Санкт-Петербург

2023

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 3](#_Toc133343243)

[1 **Захват движений человека** 4](#_Toc133343244)

[2 **Обзор существующих методов захвата движений человека** 8](#_Toc133343245)

[2.1 Захват движений с использованием маркеров 8](#_Toc133343246)

[2.1.1 Система, основанная на захвате цветных маркеров 8](#_Toc133343247)

[2.1.2 Система, основанная на маркерах с нанесенным паттерном 18](#_Toc133343248)

[2.2 Захват движений без использования маркеров 21](#_Toc133343249)

[2.2.1 Система, основанная на нейросетевых алгоритмах 21](#_Toc133343250)

[2.2.2 Системы, основанные на камерах глубины 25](#_Toc133343251)

[2.3 Выводы по данной главе 37](#_Toc133343252)

[3 **Разработка системы захвата движений на основе веб-камер** 39](#_Toc133343253)

[3.1 Устройство и принцип работы системы 39](#_Toc133343254)

[3.2 Калибровка камеры, нахождение внутренних параметров 41](#_Toc133343255)

[3.3 Калибровка пары камер, нахождение внешних параметров 43](#_Toc133343256)

[3.4 Триангуляция, нахождение трехмерных координат маркеров 46](#_Toc133343257)

[4 **Программная реализация алгоритма** 49](#_Toc133343258)

[4.1 Калибровка камер 49](#_Toc133343259)

[4.2 Калибровка пары камер 51](#_Toc133343260)

[4.3 Нахождение трехмерных координат маркеров 53](#_Toc133343261)

[4.4 Выводы по разделу 55](#_Toc133343262)

[ЗАКЛЮЧЕНИЕ 56](#_Toc133343263)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 57](#_Toc133343264)

ВВЕДЕНИЕ

Системы захвата движения имеют широкий спектр возможных сфер применения в робототехнике. Они используются для создания наиболее эргономичной рабочей среды, в которой необходимо взаимодействие робота и человека, с целью минимизации рисков возникновения травм и изучения возможности усовершенствования рабочих мест, а также повышения эффективности труда путем использования коллаборативных роботов [1]. Захват движений применяется для обучения роботов путем демонстрации в качестве основного источника входных данных для нейронных сетей [2], что увеличивает эффективность процесса обучения роботов и облегчает задачу их программирования. Управление роботами – еще одна задача, решаемая в том числе при помощи захвата движений человека, так как он предоставляет оператору наиболее естественный способ управления, что совершенствует качество взаимодействия с окружением [3], телеуправление позволяет проводить работы в агрессивной среде без непосредственного присутствия человека, а также выгодно в условиях, когда у уникальных специалистов отсутствует возможность физического пребывания в определенном месте и необходимо их дистанционное вмешательство. При этом повышение экономической доступности систем захвата движений представляется актуальной задачей, поскольку предоставляет возможности для большего их распространения, что, учитывая ряд вышеприведенных преимуществ, открывает перспективы, к примеру, для более широкого внедрения коллаборативных роботов в производственные процессы, а также упрощения процесса их обучения.

Целью работы является исследование и анализ существующих систем захвата движений человека, выявление достоинств и недостатков каждой из рассмотренных систем, а также разработка системы захвата движений на основе веб-камер массового сегмента и фидуциальных маркеров ArUco.

1 **Захват движений человека**

Захват движений – это процесс записи реального движения и его оцифровки, основанный на отслеживании определенного количества ключевых точек в пространстве с течением времени и объединении полученных координат для получения трехмерного представления записанного движения. Ключевые точки – это части, наиболее полно характеризующие движение объекта, для человека такими точками служат суставы, в которых происходит соединение частей тела и их вращение. Положение ключевых точек, помимо этого, может быть обозначено при помощи датчиков или маркеров, закрепленных на теле.

Существует большое количество систем захвата движений, классификация которых основывается на типе используемых датчиков. Данная классификация представлена на рисунке 1.1.

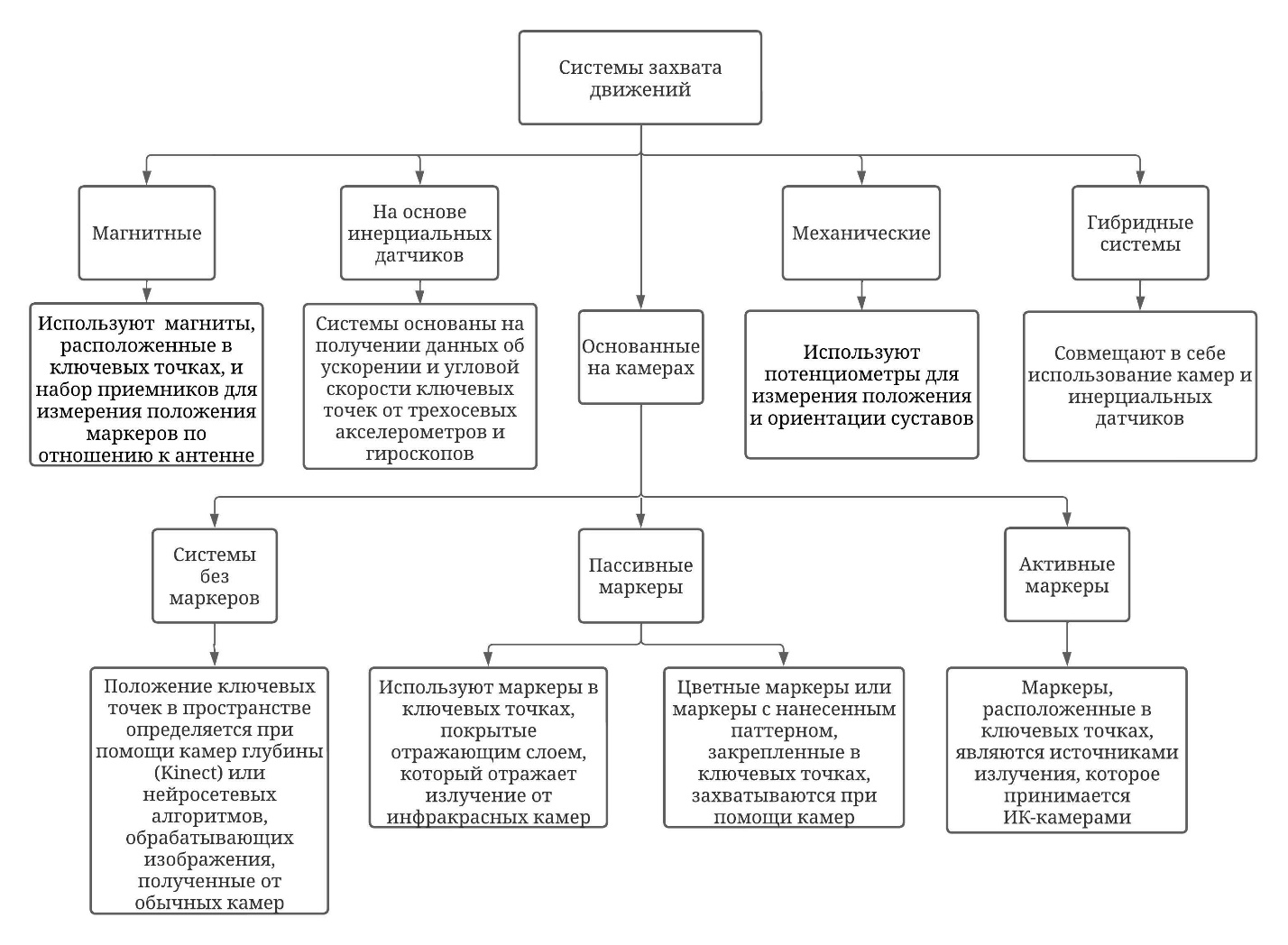


Рисунок 1.1 – Классификация систем захвата движений

Устройства, используемые для получения исходных данных о положениях ключевых точек в различных исследованиях, согласно вышеуказанному разделению систем на типы, представлены в таблице 1.1.

Таблица 1.1 – Устройства, использованные в исследованиях захвата движений

|  |  |  |  |
| --- | --- | --- | --- |
| Исследование | Тип системы | Использованные устройства | Место размещения датчиков |
| [4] | На основе камер, без использования маркеров | Kinect  ИК-камер ARTtrack2 | – |
| [5] | На основе инерциальных датчиков | Инерциальный измерительный блок | Рука |
| [6] | На основе камер, без использования маркеров | Kinect | – |
| [7] | На основе камер, без использования маркеров | Камеры глубины | – |
| [8] | Магнитная | Магнитных датчика  Излучатель | Нога |
| [9] | На основе инерциальных датчиков | Инерциальных измерительных блоков | Все тело |
| [10] | На основе камер, без использования маркеров | Камер | – |
| [11] | На основе камер, пассивные маркеры | OptiTrack V120: Trio  Маркера | Маркеры на специальном приспособлении |
| [12] | На основе камер, пассивные маркеры | ИК-камер Vicon  Маркеров | Все тело |
| [13] | Механическая | Потенциометров | Рука |
| [14] | Гибридная | Инерциальных измерительных блока  ИК-камеры Optitrack ( активных маркеров) | Верхняя часть туловища |
| [15] | На основе камер, пассивные маркеры | Камера  Цветных маркеров | Все тело |
| [16] | На основе камер, без использования маркеров | Kinect | – |
| [17] | На основе камер, пассивные маркеры | Веб-камер  Маркеров с паттерном | Нижняя часть тела |
| [18] | На основе камер, пассивные маркеры | Kinect  Цветных маркеров | Все тело |

Продолжение таблицы 1.1

|  |  |  |  |
| --- | --- | --- | --- |
| Исследование | Тип системы | Использованные устройства | Место размещения датчиков |
| [19] | На основе камер, без использования маркеров | Kinect | – |
| [20] | На основе камер, без использования маркеров | Kinect | – |

Золотым стандартом среди систем захвата движений считаются системы на основе маркеров: VICON и OptiTrack, – используемые в лабораториях биомеханики для проведения исследований человеческих движений. Однако их широкое применение ограничивается высокой ценой и необходимыми техническими навыками для сбора и анализа полученных данных.

Механические системы представляют серьезные ограничения для движений оператора в связи с жесткой структурой составных частей. К тому же возникают проблемы с определением реального положения человека, данные системы не способны измерить прыжки и повороты, а только лишь относительные углы между суставами, что делает возможным их применение только в совокупности с другими системами захвата движений.

Магнитные системы обладают ограниченной площадью действия вокруг излучателя и их переносимость также зависит от степени портативности излучателя, к тому же датчики снижают естественность движений оператора. Основным недостатком данных систем является искажение измерений из-за близости металлических объектов, а также других магнитных полей в окружающем пространстве, что снижает их применимость в различных условиях среды.

Одними из часто применяющихся систем являются системы на основе инерциальных датчиков – акселерометров и гироскопов, положения ключевых точек в которых определяются путем интегрирования полученных данных. Но интегрирование шумов в данных датчиках приводит к постепенному снижению точности определения координат и углов поворота суставов [21, 22]. Так, согласно [21] максимальный дрейф угла в инерциальном измерительном блоке доступной цены, закрепленном на кардановом подвесе с тремя степенями свободы, составлял , что значительно искажало измерения ориентации суставов.

Учитывая вышеприведенные недостатки рассмотренных систем, в данной работе решено было остановиться на рассмотрении систем, использующих камеры для захвата движений, но при этом предоставляющих наиболее экономически эффективные решения, а именно системы без применения маркеров и системы, использующие пассивные маркеры без светоотражения.

2 **Обзор существующих методов захвата движений человека**

2.1 **Захват движений с использованием маркеров**

2.1.1 Система, основанная на захвате цветных маркеров

Система VBHAS V2.0 (Video-based Human Animation System) [15] базируется на алгоритме распознавания нательных маркеров по видеоряду. Система принимает на вход видеоряд, содержащий человека, к ключевым точкам на теле которого закреплены маркеры. Результатом ее работы является выходная последовательность 3D-координат каждой ключевой точки. Далее рассматривается заложенный в ее основу алгоритм распознавания маркеров.

Очень важным является выбор правильного цветового пространства для последующей оценки цветовых диаграмм и выделения контуров. Согласно [23] выбирается наиболее эффективная система из трех ортогональных признаков, определенных по формулам (2.1) – (2.3).

где – интенсивность красного, зеленого и синего цвета от 0 до 255 соответственно.

Положение цветных маркеров на видеоряде рассматривается как динамическая система. Принимается, что – позиция точки на изображении, которая может быть представлена в виде , где - фактическая координата точки, а – Гауссовский шум. Траектория точки представляется в виде полинома третьей степени. Определяется переменная состояния , где – скорость точки, а – ее ускорение. Для предсказания положения точки на следующем кадре используется фильтр Калмана, а на рисунке 2.1 представлен результат предсказаний положения нагрудного маркера.

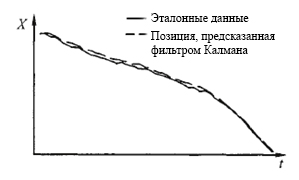


Рисунок 2.1 – Результат предсказания координат точки

Основная задача получения координат суставов для каждого кадра разбивается на две: обнаружение маркеров и их сопоставление. В данном случае применяются рамки, получаемые следующим образом: на первом кадре для получения начальных координат суставов применяется статическая модель, на основе которой рамка размером определяется для каждого маркера, и, следовательно, последующие алгоритмы отслеживания ограничиваются данной рамкой. На последующих кадрах за координаты центра рамки принимаются координаты, полученные фильтром Калмана. Размер рамки может динамически изменяться для охвата всего маркера.

Для определения контура маркера применяется алгоритм Кэнни [24]. Данный метод применяется для изображений, приведенных к оттенкам серого, поэтому в данном случае метод Кэнни применяется к трем изображениям, представляющим собой три ортогональные составляющие, представленные выше, а затем применяется логическое ИЛИ для получения окончательных границ. После расчета градиентов алгоритм использует две пороговые величины: верхнюю - и нижнюю - – для выделения контуров. Правила для каждого пикселя таковы: (1) если градиент больше , то это пиксель, принадлежащий границе, (2) если градиент меньше , то данный пиксель не имеет отношения к границе, (3) если же градиент принимает промежуточное значение, то применяется следующее дополнительное правило: если от данного пикселя до какого-либо пикселя границы может быть найден путь, на котором значения градиента для каждого пикселя превышает значение , то все пиксели, входящие в данный путь, помечаются как граничные.

Для повышения эффективности алгоритма с учетом того, что детекция производится на последовательных кадрах, пороговые значения динамически изменяются для отражения межкадровой корреляции. В местах, где границы были обнаружены на предыдущих кадрах, данные значения понижаются для повышения вероятности обнаружения граничных пикселей, а в местах отсутствия границ значения повышаются для снижения возможных шумов. На рисунке 2.2 представлено сравнение традиционного и улучшенного алгоритмов.



Рисунок 2.2 – Сравнение методов Кэнни

Границы на полученных бинарных изображениях не являются непрерывными, следовательно, не формируют необходимых замкнутых контуров. Для решения данной проблемы применяется преобразование Хафа, которое основывается на представлении искомого объекта в виде параметрического уравнения и используется для поиска простых форм на изображении, в данном случае для детекции прямоугольных маркеров. Прямая для преобразования Хафа представляется, согласно уравнению (2.4), следующим образом:

где – длина нормали к прямой, проведенной из начала координат;

– угол между нормалью и осью абсцисс.

Результат применения преобразования Хафа представлен на рисунке 2.3.

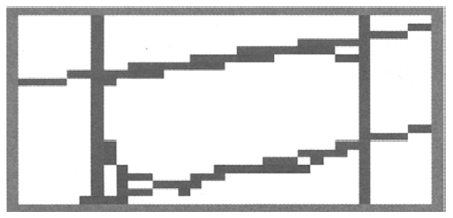


Рисунок 2.3 – Преобразование Хафа

Шумы на изображении, а также возможные искажения формы маркеров усложняют задачу получения необходимых четырех линий, формирующих прямоугольник. После преобразования Хафа получены восемь линий с наибольшими значениями в аккумуляторе Хафа, для выбора четырех из них используются два следующих алгоритма.

Первый алгоритм основан на анализе границ. Каждая прямая , полученная из преобразования Хафа может быть представлена в виде уравнения (2.5):

К тому же каждая пара прямых связана одним из двух соотношений: параллельностью или перпендикулярностью. Допустим, что – прямая с наибольшим значением в аккумуляторе, для нее необходимо найти прямую, параллельную ей, и две прямых, ей перпендикулярных. Условие параллельности задается уравнениями (2.6), (2.7):

где – пороговое значение угла, которое соответствует условию параллельности прямых (;

– пороговое значение расстояния, которое введено для проверки несовпадения двух прямых.

Условие перпендикулярности задается уравнением (2.8):

Для снижения влияния шумов формирование прямоугольников происходит в порядке приоритета, который задается линиям согласно их значениям в аккумуляторе. Возможны ситуации, когда одна или более граница оказываются потерянными, в таком случае, если потеряно больше одной прямой, для формирования прямоугольника используется второй алгоритм, если же не найдена лишь одна граница, она может быть восстановлена в рамках данного алгоритма. При известной уравнении прямой и координатах центра маркера , полученных из фильтра Калмана, определяется уравнение (2.9) прямой .

Так как и симметричны относительно , получаем уравнение (2.10):

Следовательно, получим уравнение (2.11):

Алгоритм 2. Формирование прямоугольника по предыдущему кадру.

Из предыдущего кадра получено четыре прямых . Тогда в данном кадре из восьми прямых с наибольшими значениями выбираются четыре по критерию минимизации целевой функции , заданному уравнением (2.12):

где – весовые коэффициенты, причем , так как оказывает большее влияние.

Алгоритмы применяются в следующем порядке. Первый алгоритм рассматривает лишь состояние в данном кадре, не учитывая предыдущих. Он способен восстановить потерянную линию, но это может привести к появлению ошибочных прямых. Следовательно, результат работы данного алгоритма должен быть проверен с помощью использования информации о цвете и форме. Если проверка не удается, то выполняется второй алгоритм, который использует междкадровую информацию и в большинстве случаев устойчив к ошибкам. Его недостаток заключается в том, что ошибки из предыдущих кадров будут распространяться на последующие кадры. Поэтому в данном случае второй алгоритм используется лишь в качестве дополнения к первому.

Для получения 3D информации о движении человека по 2D данным используются стартовые точки, от которых затем отталкиваются для получения информации обо всех суставах. Прежде всего для получения стартовых точек проводится калибровка камеры.

Перспективной проекции точки в мировых координатах на плоскость соответствует точка , координаты которой определяются по уравнениям (2.13) и (2.14):

где – координаты камеры,

– матрица поворота,

– матрица перехода,

– фокусное расстояние.

, и могут быть определены заданием известных точек в пространстве и их проекций. Пока пропорции отслеживаемых объектов соответствуют действительности, положение и ориентация камеры могут быть определены, следовательно, найдены , и , что гарантирует схождение определенных 3D и 2D точек.

В качестве начальных точек используется треугольник «грудь-левое плечо-правое плечо», так как данные точки меньше всего подвержены перекрыванию и деформации.

Если координаты точек стартового треугольника на *n*-ом кадре равны и скорость каждой точки равна , следовательно интервал поиска для кадра может быть определен как . Учитывая, что изменение координат треугольника происходит постепенно, можно ограничить интервал значений координат и следующим образом, согласно уравнениям (2.15) и (2.16):

где

.

После получения координат стартового треугольника необходимо перенести результаты на все суставы. В перспективной проекции точка на плоскости соответствует прямой в 3D системе координат, поэтому для того, чтобы определить положение на этой прямой, необходимо использовать дополнительную информацию о модели человека или размерах скелета. По известным координатам проекции может быть найдена прямая, на которой лежат точки, расстояние от которой до смежной вершины (стартового сустава) с известными координатами, будет соответствовать реальному размеру сустава. Проблема заключается в том, что, как показано на рисунке 2.4, таких точек может быть две.

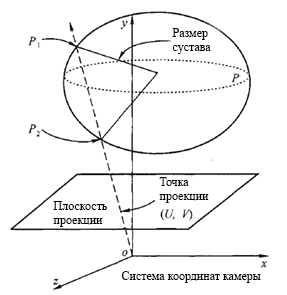


Рисунок 2.4 – Неоднозначность полученных результатов

Для решения данной проблемы применяется анализ 2D признаков, находятся кадры, на которых происходят резкие, значительные изменения и, таким образом, видео разбивается на сегменты. На рисунке 2.5 приведены значения скоростей суставов руки для тридцатитрехкадрового клипа, на котором запечатлен момент прыжка.

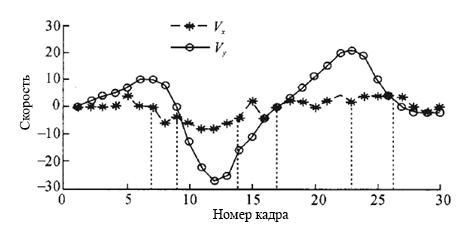


Рисунок 2.5 – График движения руки

На данном графике можно обнаружить шесть точек, в которых какая-либо из скоростей принимает нулевое значение. После анализа самого клипа становится ясно, что данные нулевые точки соответствуют точкам значительных изменений на кадре, следовательно, на их базе и производится сегментация видеоряда. Для каждого сегмента , на котором не происходит резких изменений движения, проблема неоднозначности может быть решена в три шага.

Первый шаг. Два положения удовлетворяют известному изображению. Так как в используемом стартовом треугольнике три вершины, структура данных порождаемая поиском необходимой точки представляет из себя деревьев.

Второй шаг. Часть ветвей может быть отброшена на основании имеющихся данных. Например, на коленный, бедренный и голеностопный суставы наложено ограничение, согласно которому первый из перечисленных всегда располагается впереди остальных двух. Помимо этого, пользователь способен сам накладывать ограничения в зависимости от типа рассматриваемого движения.

Третий шаг. Если два предыдущих шага не решают указанной проблемы, то на последнем шаге используется предположение о непрерывности движения. Данный метод рассматривается на примере дерева, стартовой точкой которого является левое плечо, и состоит оно из локтевого и кистевого суставов, при этом рассматривается лишь одна степень свободы. Определяется целевая функция для кадров , и согласно уравнению (2.17):

где – угловое ускорение на трех кадрах.

На основе предыдущего уравнения получено уравнение (2.18):

где - функция, определяющая степень сглаживания в пределах кадров.

Таким образом, для каждого кадра строится граф, представленный на рисунке 2.6, включающий в себя начальную точку и конечную точку . Таким образом, поиск удовлетворяющих вершин преобразуется в решение проблемы поиска кратчайшего пути в графе.

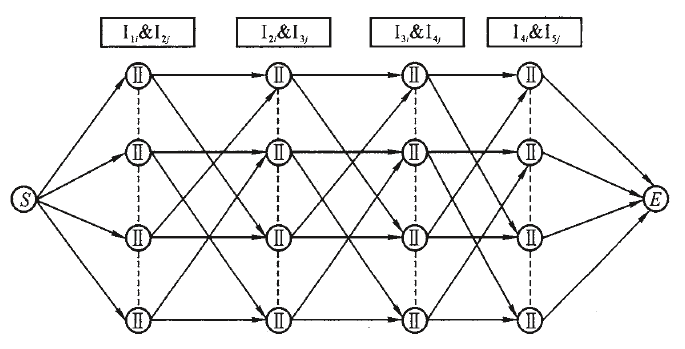


Рисунок 2.6 – Граф точек

2.1.2 Система, основанная на маркерах с нанесенным паттерном

Данная система [17] состоит из пяти пар веб-камер, расположенных по периметру снимаемой сцены, которая включает в себя человека, на ключевых точках тела которого закреплены маркеры с уникальным паттерном. Когда на изображениях, полученных с двух или более камер, в результате работы алгоритма распознавания обнаруживается одинаковый паттерн, производится определение 3D-координат соответствующей ключевой точки с помощью метода мультилатерации, то есть нахождения координат точки по известному расстоянию от нее до более чем двух точек. Далее описывается лежащий в основе системы алгоритм распознавания маркеров.

Паттерн представляет собой нанесенную черным цветом поверх белого фона последовательность из двух объединенных шестиугольников, разделенных на треугольники, в части из которых отмечены точки. Каждый шестиугольник может быть определен путем считывания по часовой стрелке последовательности треугольников с точкой и без нее (треугольникам с точкой внутри соответствует единица, а пустым – ноль), при этом при формировании шестиугольников исключалось использование последовательностей с радиальной симметрией. Используемые в данном методе шестиугольники представлены на рисунке 2.7.

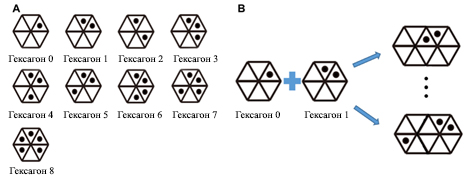


Рисунок 2.7 – Используемые шестиугольники

Первым шагом в процессе распознавания маркеров служит пороговая обработка изображения, позволяющая получить бинарные кадры, на базе которых осуществляется распознавание контура паттерна и треугольников, его составляющих. Для выделения треугольников применяется алгоритм Рамера–Дугласа–Пекера, позволяющий упростить ломанную линию путем уменьшения количества составляющих ее точек. Распознавание шестиугольников осуществляется с помощью областей интереса, представляющих собой квадраты с длиной стороны, равной длине стороны треугольника, координаты центра которых соответствуют координатам вершин каждого треугольника. Когда внутри области оказывается одновременно шесть вершин, двоичный код, соответствующий каждому шестиугольнику, проверяется по часовой стрелке, начиная с треугольника, вершина которого служит центром данной области. На последнем шаге путем определения наложенных друг на друга треугольников из двух разных шестиугольников, составляющих последовательность, производится распознавание окончательного паттерна, который формируется из номеров шестиугольников и номеров перекрывающих друг друга треугольников. Процесс распознавания маркеров представлен на рисунке 2.8.

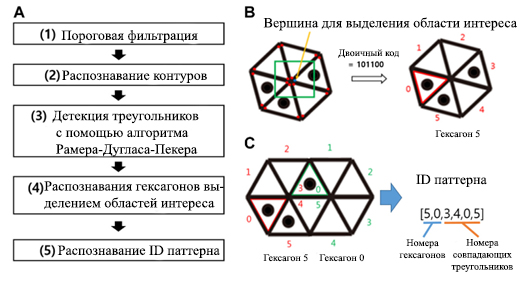


Рисунок 2.8 – Распознавание маркеров

Процесс оценки положения тела человека состоит из двух этапов: анализа статичного положения тела и анализа тела в движении. На первом шаге, представленном на рисунке 2.9, определяются геометрические параметры частей тела, на которые надеты маркеры. Для определения направления осей на каждом из маркеров предварительно выбираются передний и задний шестиугольники, центральные точки которых определяются, соответственно, как передняя и задняя точки той части тела, на которой размещен маркер. Положительному направлению оси Y соответствует вектор, направленный из задней точки к передней, положительному направлению оси Z соответствует вектор, нормальный к плоскости, проходящей через центральные точки всех шестиугольников. Начало координат определяется как точка, лежащая по середине отрезка между задней и передней центральными точками. Далее определяются расстояние от начала координат до каждой вершины паттерна, а также угол между направлением оси Y и вектором, проведенным от начала координат к центру каждого шестиугольника.

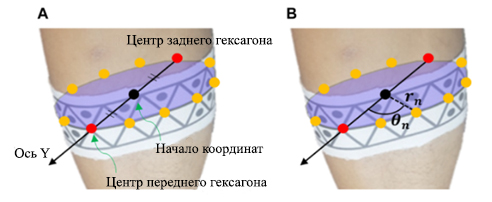


Рисунок 2.9 – Определение направления осей

Второй этап включает в себя анализ в движении, по определенным на изображении центрам шестиугольников и по известным расстояниям от них определяется положение центра координат данной части тела. Направление оси Z задается вектором, направленным по нормали к плоскости, которая проходит через точки центров и найденный центр координат. Направление оси Y находится путем поворота вектора, соединяющего центр координат с одним из захваченных центров шестиугольника, на величину угла, определенную на этапе статической оценки. Определение положения начала координат осуществляется с помощью мультилатерации, то есть нахождения 3D-координат точки по известному расстоянию от нее до более чем двух точек.

Преимущества данных методов:

– Большая по сравнению с системами без маркеров точность [25, 26];

– Захват движений может быть реализован на основе камер с низкой стоимостью [15, 17];

– 3D-координаты могут быть получены с использованием только одной камеры [15];

Недостатки:

– Артефакты, вызванные движением мягких тканей, отделяющих маркер от кости, известные также как STA (Soft Tissue Artifacts) [27, 28];

– Снижение точности измерений из-за возможного перекрытия маркеров [29, 30]:

– Наличие маркеров может привести к снижению естественности движений оператора;

– Подготовка системы к работе требует продолжительного времени в связи с необходимостью точной установки маркеров;

**2.2 Захват движений без использования маркеров**

2.2.1 Система, основанная на нейросетевых алгоритмах

Данная система [10] состоит из девяти камер, расположенных по периметру снимаемой сцены, для получения координат ключевых точек изображения со всех камер обрабатывались при помощи различных систем: *OpenPose*, *AlphaPose* и *DeepLabCut*, - позволяющих определять двумерные координаты ключевых точек на изображении, далее будет рассмотрена система *OpenPose*, результаты которой оказались наиболее близки к результатам эталонной системы захвата движений - *Qualysis*. Для получения 3D-координат для каждого последовательного обнаружения сустава был применен алгоритм обратного проецирования, результатом которого является набор лучей в пространстве, точка пересечения которых есть центр соответствующего сустава. Далее приводится описание системы, лежащей в основе получения двумерных координат ключевых точек.

На вход рассматриваемой системы [31, 32] подается изображение размером , в качестве результата выдаются 2D координаты ключевых точек тела каждого человека, изображенного на фотографии. В первую очередь при помощи нейронной сети прямого распространения создается набор карт достоверности (*Confidence maps*) и полей сходства частей тела (*Part Affinity Fields*). Затем для получения координат ключевых точек карты достоверности и поля сходства анализируются при помощи жадного алгоритма.

Карта достоверности – это матрица, которая содержит вероятность нахождения каждом пикселе определенной ключевой точки, то есть, например, если на фотографии исходного размера необходимо найти 20 ключевых точек, будет создано 20 карт достоверности размером , каждая клетка которых будет содержать значение вероятности нахождения в данном пикселе искомой ключевой точки.

Поле сходства частей тела – векторное поле, которое характеризует степень связанности ключевых точек (принадлежности их одному человеку, например, в случае нахождения в кадре нескольких человек). Каждый вектор задает направление конечности, соединяющей пару ключевых точек. Пример карты достоверности и поля сходства представлен на рисунке 2.10.

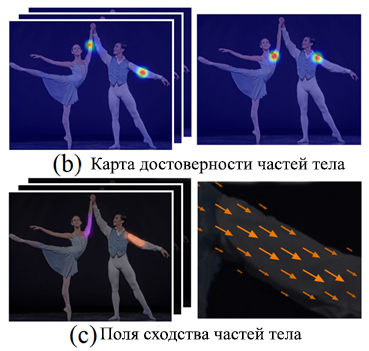


Рисунок 2.10 – Карта достоверности и поле сходства

Для получения карты признаков, являющейся входными данными данной архитектуры, используются первые десять слоев сверточной нейронной сети VGG-19, структура которой представлена на рисунке 2.11.

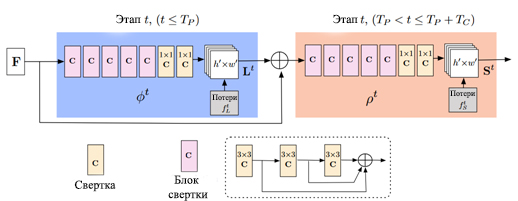


Рисунок 2.11 – Структура сверточной нейронной сети

На данном этапе нейронная сеть создает набор полей сходства частей тела согласно уравнению (2.19)

где – вывод сверточной нейронной сети на первом этапе.

Для усовершенствования предсказаний на каждой последующей стадии производится конкатенация результатов предыдущей стадии и карты признаков исходной фотографии, заданных уравнением (2.20):

где - вывод сверточной нейронной сети на этапе ,

– общее количество стадий формирования полей сходства.

Данный процесс по завершении итераций повторяется для создания карт достоверности (данный этап выделен бежевым цветом), в качестве входных данных используется последнее уточненное поле сходства частей тела согласно уравнениям (2.21) и (2.22):

где - вывод сверточной нейронной сети на этапе ;

– общее количество стадий создания карт достоверности.

Для настройки нейросети во время итерационного получения карт достоверности и полей сходства применяется функция потерь, которая задает потери между полученными данными и эталонными значениями карт и полей. Данные функции для этапа создания полей сходства и этапа создания карт достоверности задаются уравнениями (2.23) и (2.24):

где и - эталонное поле сходства и карта достоверности,

– двоичная маска, которая используется для предотвращения потери верных результатов.

Для решения проблемы затухающего градиента применяется промежуточный контроль, описанный в [33], увеличивающий значение градиента после каждой стадии. Итоговая функция потерь задается уравнением (2.25):

Для получения 3D координат человеческого тела с помощью данного метода необходимо одновременное использование двух или более камер с последующим сопоставлением данных, полученных с их помощью.

Преимущества системы:

– Отсутствие маркеров не ограничивает движений оператора

– Настройка системы не требует продолжительного времени на закрепление маркеров

Недостатки:

– Большинство алгоритмов рассматривают последовательные временные интервалы как статистически независимые, что приводит к ошибкам оценки положения тела в последовательных кадрах [34];

– Точность данных методов уступает системам, основанным на маркерах [17, 25, 26];

– Наибольшие ошибки в измерениях наблюдаются вследствие неправильной разметки эталонных данных на тренировочных датасетах [34, 35].

2.2.2 Системы, основанные на камерах глубины

В данных системах получение исходных данных о трехмерных координатах ключевых точек основано на применении камер глубины, лидерами среди которых на рынке являются два вида устройств: Kinect и Intel RealSense. Далее будут рассмотрены принципы действия двух поколений Kinect и семейства камер D400 Intel RealSense, так как данное поколение камер использует отличный от Kinect принцип определения глубины на изображении.

Kinect состоит из RGB-камеры, а также инфракрасных излучателя и приемника. Получение расстояний до объекта производится методом триангуляции согласно [36]. Излучатель испускает луч света, который разбивается дифракционной решеткой на множество лучей для создания постоянного паттерна из пятен, который затем захватывается приемником и соотносится с эталонным паттерном, сохраненным в памяти сенсора и полученным путем предварительного захвата плоскости, находящейся на известном расстоянии. Когда дифракционное пятно проецируется на объект, расстояние от которого до камеры меньше или больше расстояния от камеры до эталонной плоскости, проекция пятна на матрицу инфракрасной камеры будет смещена в направлении прямой, соединяющей излучатель с перспективным центром приемника, относительно проекции, полученной от эталонной плоскости. Данные смещения определяются для всех пятен путем соотнесения полученного изображения с эталонным и именуются несоответствиями или диспаратностями. С помощью данных несоответствий затем вычисляется расстояние до каждого пикселя.

Рисунок 2.12 отражает соотношение между расстоянием от камеры до точки объекта , расстоянием до эталонной плоскости и полученным смещением .

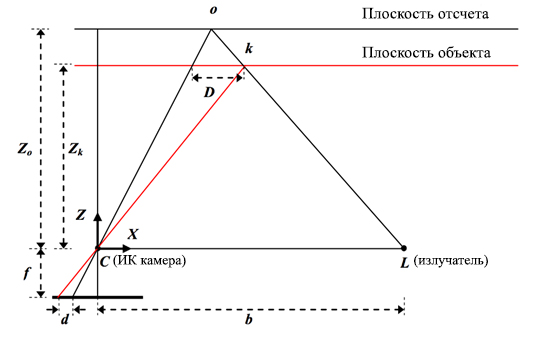


Рисунок 2.12 – Схема получения расстояния до объекта

Для получения трехмерных координат начало системы координат располагается в перспективном центре инфракрасной камеры. Ось *Z* ортогональна плоскости изображения и направлена от камеры к объекту, ось *X* перпендикулярна оси *Z* и направлена вдоль базовой линии *b* в направлении излучателя, а ось *Y* перпендикулярна предыдущим осям согласно правоориентированному базису. Предположим, что – расстояние до эталонной плоскости, тогда, если объект расположен ближе, то проекция пятна на объекте будет находиться левее проекции от эталонной плоскости, что выражается смещением . Тогда из подобия треугольников получим уравнения (2.26), (2.27):

где – расстояние до объекта,

– расстояние между излучателем и приемником,

– фокусное расстояние инфракрасной камеры,

– сдвиг точки в плоскости объекта,

– смещение в плоскости изображения.

Выразив из второго уравнения и подставив его в первое, получим уравнение (2.28):

Уравнение является базовым для получения расстояния до объекта по известному значению диспарантности , при этом параметры и определяются в процессе калибровки. Координаты и могут быть получены из уравнений (2.29) и (2.30):

где – координаты точки на изображении,

– координаты точки пересечения плоскости изображения с оптической осью.

– поправки на искажения, вносимые линзой.

Система Kinect One представляет из себя времяпролетную камеру, технология измерения глубины которой заключается в следующем согласно [37]. Излучатель испускает определенную модулированную волну. Полученная приемником волна после отражения имеет некоторый сдвиг по фазе . Допустим, что излучается сигнал, заданный уравнением (2.31):

где – частота модуляции.

Следовательно, на приемнике будет сигнал, заданный уравнением (2.32):

где – постоянный сдвиг,

– амплитуда.

Демодуляция сигнала производится путем взаимной корреляции данных сигналов. Корреляционная функция задается уравнениями (2.33) и (2.34):

где – корелляционная функция.

Для определения на одном периоде сигнал рассматривается в четырех точках, сдвинутых друг относительно друга по фазе на , каждое из этих значений определяется из формулы (2.35):

Следовательно, сдвиг по фазе определяется согласно уравнению (2.36):

Следовательно, расстояние до объекта согласно уравнению (2.37) равно:

где – скорость света.

В основе стереоскопии лежит способность человеческого мозга оценивать глубину по двум изображениям от глаз [38]. Нахождение глубины происходит на основании несоответствия двух кадров одной сцены с двух разных точек. На рисунке 2.13 представлена базовая стереоскопическая система.

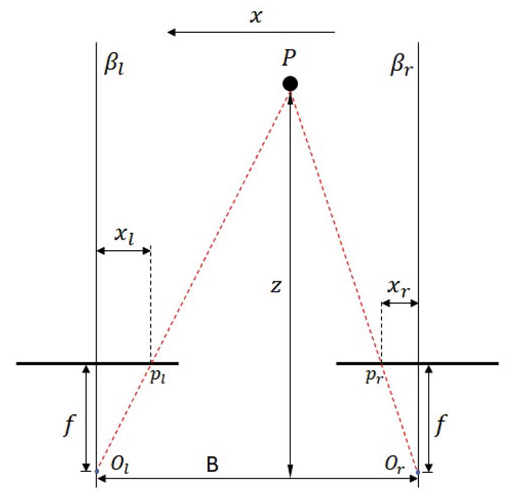


Рисунок 2.13 – Основы стереоскопии

Рисунок выше иллюстрирует упрощенную модель для триангуляции. Две камеры имеют параллельные оптические оси и одновременно снимают точку , расположенную на расстоянии . и соответствующие проекции точки на плоскость камер, а и – смещение этих точек относительно оптических осей. Тогда, зная внутренние и внешние параметры камер, из подобия треугольников и получим уравнение (2.38).

где – длина базовой линии,

– фокусное расстояние камер.

Ошибка в определении глубины растет пропорционально квадрату глубины и ошибки в смещении , которое задается разностью в знаменателе, что отражено в уравнении (2.39).

Активная стереоскопия относится к алгоритмам корреляционного сравнения, в которых производится сопоставление двух кадров попиксельно для последующего определения положения соответствующих пикселей путем триангуляции. Данный метод требует, чтобы на поверхности снимаемого объекта просматривалась четкая текстура, для повышения качества корреляции. На рисунке 2.14 представлены два облака точек, полученных при съемке одного и того же объекта, на поверхность которого во втором случае нанесена текстура.

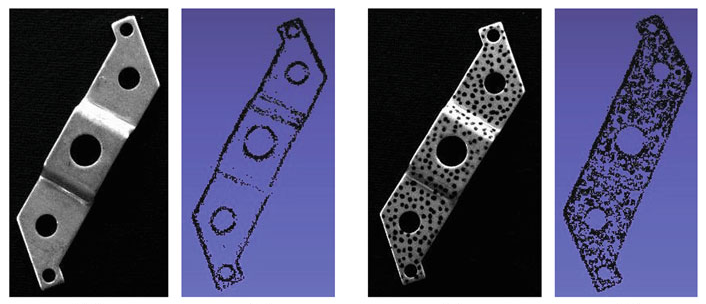


Рисунок 2.14 – Влияние текстуры на облака точек

Для решения подобной проблемы в камерах, использующих данный метод, происходит проецирование случайного светового паттерна на поверхность захватываемого объекта для упрощения процесса триангуляции. Пример такого паттерна представлен на рисунке 2.15.

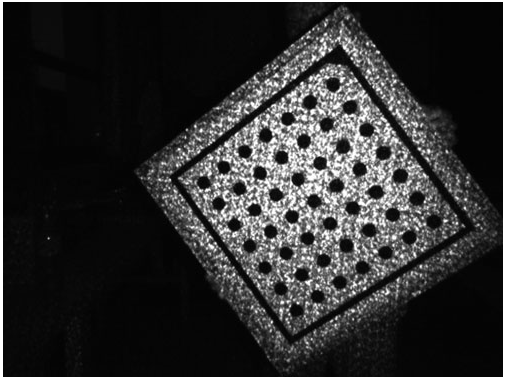


Рисунок 2.15 – Пример светового паттерна

Геометрия данных паттернов оказывает значительное влияние на качество определения глубины, так как правильный ее выбор позволяет избежать неоднозначности при сопоставлении совпадающих точек во время сравнения двух кадров. Наиболее теоретически эффективным считается паттерн из белого шума, когда интенсивность каждого пикселя не зависит от окружающих, так как это сделало бы его более плотным, что гарантирует лучшее нахождение необходимых пикселей. На практике же используются случайные паттерны в инфракрасном спектре во избежание возможного влияния окружающего света.

Для распознавания положения тела полученное с камеры глубины изображение человека разбивается на цветные сегменты [39], представленные на рисунке 2.16, часть из которых предназначена для непосредственной локализации суставов, другие же заполняют пустые промежутки или же могут использоваться совместно с другим для предсказания положения определенных суставов.



Рисунок 2.16 – Сегментация изображений

Данное представление преобразует задачу определения положения суставов в задачу, которую можно решить эффективным алгоритмом классификации. Применяются наиболее простые классификационные признаки, основанные на глубине изображения, для заданного пикселя признак задается уравнением (2.40):

где – глубина в данном пикселе на изображении ;

– параметры, описывающие смещения.

На рисунке 2.17 представлены два признака при разных положениях пикселя . Признак направлен вверх, следовательно, значения уравнения для пикселей, расположенных в верхней части тела будут большими, в то время как для пикселей ниже значения будут близки к нулю. Признак используется для нахождения тонких вертикальных частей тела, таких как рука.

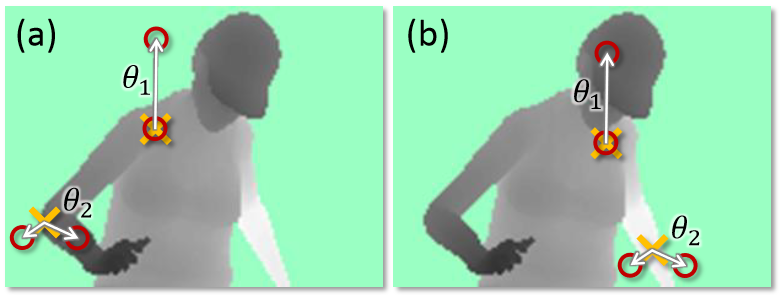


Рисунок 2.17 – Расчет признаков

Для классификации используется метод случайного леса, представляющий собой ансамбль деревьев принятия решений. На рисунке 2.18 представлен лес, состоящий из деревьев, каждое из которых включает в себя узловые и краевые вершины. Каждая узловая вершина состоит из признака и порогового значения .

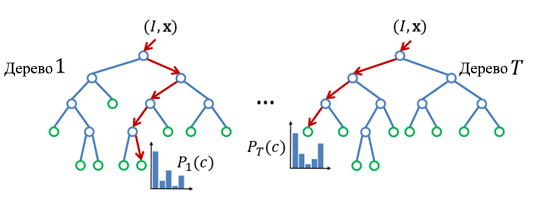


Рисунок 2.18 – Метод случайного леса

Для классифицирования пикселя последовательно рассчитывается значение по уравнению , которое сравнивается с пороговым значением для выбора следующей ветви. Полученное распределение по меткам частей тела сохраняется в листовом узле дерева . Для получения окончательной классификации берется среднее от всех распределений по всем деревьям согласно уравнению (2.41):

Распознавание частей тела использует информацию о каждом пикселе. Теперь данные о каждом пикселе должны быть объединены для получения надежных предположений о положении суставов. Для оценки плотности распределения вероятности применяется поиск моды, основанный на сдвиге среднего значения со взвешенным Гауссовым ядром, заданный уравнением (2.42):

где – положение в мировых координатах;

– количество пикселей в изображении,

– вес пикселя,

– проекция пикселя изображения в мировых координатах,

– пропускная способность.

Вес каждого пикселя, заданный уравнением (2.43), учитывает распределение для каждого пикселя и площадь поверхности, заключенной в нем:

Исследования, целью которых являлось сравнение камер Kinect и Intel RealSense, демонстрировали, что ни одна из систем не превосходит конкурента полностью [40]. В то время как первое поколение системы Kinect демонстрировало однозначно худшие результаты, Kinect v2 и Intel RealSense D400 обладали различными преимуществами. Так, согласно [41], камеры Intel обладают большей дальностью захвата, доходящей до 7 метров, в то время как максимальная дальность эффективной работы Kinect v2 составляет 4,5 метра. Однако Kinect v2 обладает меньшей погрешностью и систематической ошибкой измерений [38], хотя более уязвима к солнечному свету [42], что значительно снижает точность измерений вне помещений, в отличие от Intel RealSense. Применение данных камер для анализа человеческих движений в помещении демонстрирует меньшую уязвимость к шумам и большую точность измерений при захвате ключевых точек для Kinect v2 [43].

Преимущества систем:

– Камера глубины позволяет использовать только одно устройство для получения необходимых координат

– Отсутствие маркеров не ограничивает естественности движений оператора

– Компактность системы, удобство использования за счет наличия оригинального ПО и плагинов.

Недостатки:

– Данные подвержены значительному фазовому дрожанию и влиянию шумов [43, 44];

– Снижение точности измерений при перекрытии конечностей или выходе из поля зрения [45, 46];

– Освещенность снимаемой сцены значительно влияет на точность полученных данных [47] (Не применимо к Intel RealSense);

– Высокая стоимость новых моделей, поставки и старых, и новых прекращены.

2.3 **Выводы по данной главе**

В данной главе были рассмотрены существующие системы на основе камер, применяемые для захвата движений человека, проанализированы их достоинства и недостатки, представленные в таблице 2.1.

Таблица 2.1 – Сравнительная таблица методов захвата движений

|  |  |  |
| --- | --- | --- |
| Метод | Преимущества | Недостатки |
| На основе сверточных нейронных сетей (OpenPose, AlphaPose и др.) | Естественность движений не ограничивается за счет нательных маркеров  Упрощается процесс подготовки системы к работе | Меньшая точность из-за неправильной разметки эталонных данных в тренировочных датасетах [25, 26, 34, 35];  Последовательные временные интервалы рассматриваются как статистически независимые, что приводит к ошибкам оценки положения тела в последовательных кадрах [34];  Малое количество систем в открытом доступе |
| Камеры глубины(Kinect, Intel RealSense.) | Одного устройства достаточно для получения координат  Не ограничивает движений оператора  Простота использования системы | Высокая уязвимость к шумам [43]  Снижение точности измерений при перекрытии конечностей или выходе из поля зрения [45, 46]  Значительное снижение точности в зависимости от освещенности сцены [44, 47] (Не применимо к Intel Realsense)  Данные подвержены значительному фазовому дрожанию [44];  Высокая стоимость, прекращены поставки |

Продолжение таблицы 2.1

|  |  |  |
| --- | --- | --- |
| Метод | Преимущества | Недостатки |
| На основе маркеров | Большая по сравнению с предыдущими системами точность [25, 26]  Могут быть реализованы на основе камер с низкой стоимостью [15, 17]  3D-координаты могут быть получены с использованием только одной камеры [15] | Снижение точности измерений в случае перекрытия маркеров [29, 30]  Возникновение артефактов, вызванных движением мягких тканей, отделяющих маркер от кости (Soft Tissue Artifacts) [27, 28]  Более длительная настройка системы, расположение маркеров в нужных местах  Возможное снижение естественности движений оператора |

Системы, основанные на нательных маркерах, все еще остаются наиболее надежными с точки зрения точности, несмотря на сопутствующие проблемы, связанные с ухудшением качества измерений в связи движением мягких тканей, возможным перекрытием маркеров, а также ограничением естественности движений оператора.

3 **Математическое описание алгоритма захвата движений**

Описываемый алгоритм основан на системе, состоящей из двух камер, с помощью которых производится захват маркеров, расположенных в ключевых точках тела. Данные камеры образуют стереосистему, зная параметры которой, можно по координатам точки на полученных изображениях определить положение точки в пространстве при помощи триангуляции. При этом связь координат точки на изображении и в пространстве можно обозначить при помощи проективной модели камеры, представленной на рисунке 3.1.

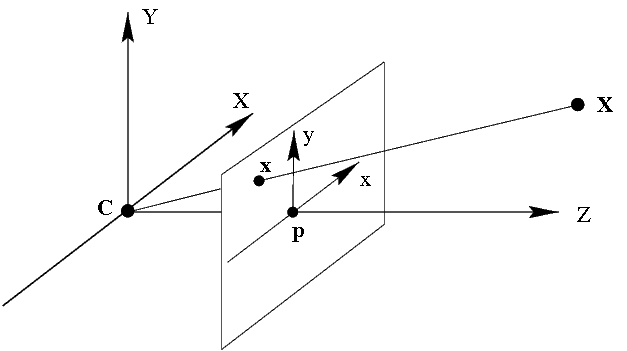


Рисунок 3.1 – Модель проективной камеры

На рисунке выше точка в пространстве проецируется на плоскость изображения в точку , при этом точка – оптический центр камеры, а луч – оптическая ось. При использовании данной модели камеры формула, задающая проецирование точки на плоскость изображения, принимает следующий вид в однородных координатах (3.1):

где – однородные координаты точки на изображении,

– однородные координаты точки в пространстве,

– матрица камеры.

Матрица камеры задает переход из мировой системы координат в систему координат камеры при условии, что центр камеры не совпадает с началом координат системы, а также переход из системы координат камеры в систему координат изображения. Данная матрица задается уравнением (3.2):

где – матрица внутренних параметров камеры,

– матрица поворота системы координат камеры относительно мировой системы координат,

– вектор перемещения системы координат камеры относительно мировой системы координат.

Следовательно, для нахождения трехмерных координат маркеров необходимо определить матрицу камеры, а точнее ее компоненты: внутренние и внешние параметры камеры, нахождение которых осуществляется при помощи калибровки.

3.1 **Калибровка камеры, нахождение внутренних параметров**

Матрица внутренних параметров камеры задает переход из системы координат камеры в систему координат изображения. В простейшем случае матрица имеет следующий вид (3.3):

где – фокусное расстояние камеры.

Однако начало координат изображения может не совпадать с оптическим центром камеры, будучи смещенным на определенное расстояние. К тому же в случае с CCD камерами существует вероятность того, что ячейки матрицы будут не квадратными, а прямоугольными, что вносит неодинаковые масштабные коэффициенты по осям, то есть, если количество пикселей на единицу длины по одной оси составляет , а по другой – , то соответствующие фокусные расстояния будут задаваться как , что с учетом несовпадения оптического центра с началом координат изображения приводит матрицу внутренних параметров к следующему виду (3.4):

где – координаты оптического центра камеры в системе координат изображения;

– параметр, задающий перекос осей друг относительно друга, обычно, в том числе и в OpenCV, принимается равным нулю.

Помимо нахождения матрицы внутренних параметров камеры, необходимо учесть, что оптические системы – линзы – подвержены возникновению аберраций, в данном случае, радиальной и тангенциальной дисторсиям. Радиальная дисторсия возникает вследствие неравномерности преломления света из-за сферической поверхности линзы объектива, так как лучи сильнее преломляются ближе к краям линзы, что искажает прямые линии. Тангенциальная дисторсия возникает из-за наклона плоскости матрицы сенсора по отношению к линзе. Уравнения дисторсии согласно математической модели Брауна[48] задаются следующим образом для радиальной и тангенциальной дисторсий (3.5), (3.6):

где – положение точки после коррекции дисторсии;

– коэффициенты радиальной дисторсии;

*–* расстояние до центра изображения.

– коэффициенты тангенциальной дисторсии.

3.2 **Калибровка пары камер, нахождение внешних параметров**

После получения внутренних параметров камер необходимо определить внешние параметры, которые задают положение в пространстве одной камеры относительно другой. Определение матриц поворота и перемещения основано на ограничениях, накладываемых эпиполярной геометрией на систему из двух камер, обобщенная схема которой с указанием основных элементов эпиполярной геометрии представлена на рисунке 3.2.

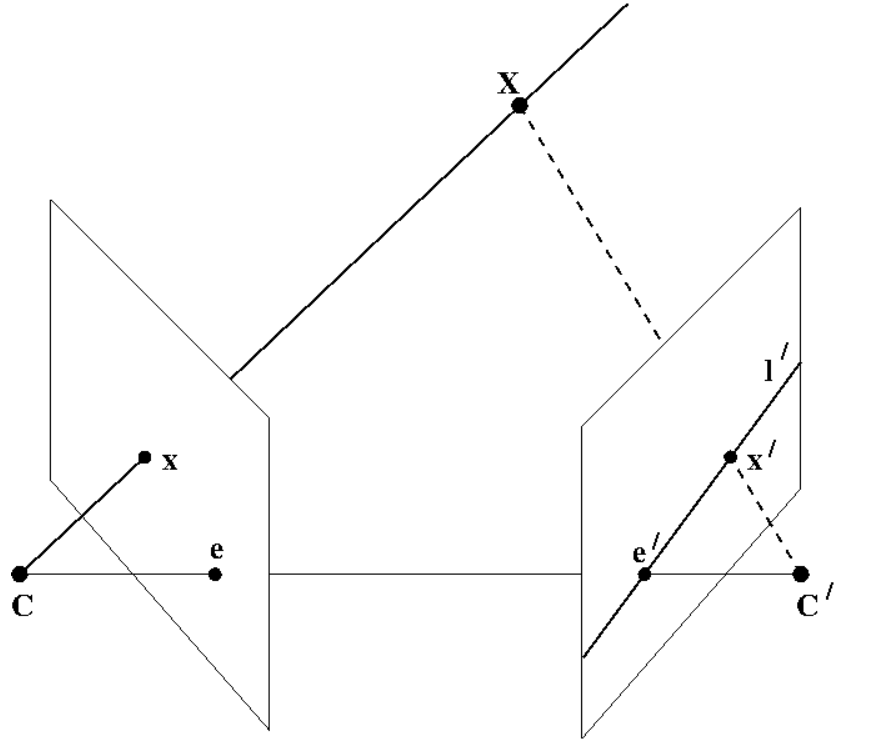


Рисунок 3.2 – Эпиполярная геометрия

Точка проецируется в точку на изображении левой камеры и в точку на плоскости изображения правой камеры. Точка лежит на эпиполярной линии , которая является проекцией луча на плоскость изображения правой камеры и образована пересечением этой плоскости с эпиполярной плоскостью. Точки и – центры камер, а и - эпиполи. Обозначим вектор как , а как согласно рисунку 3.3.

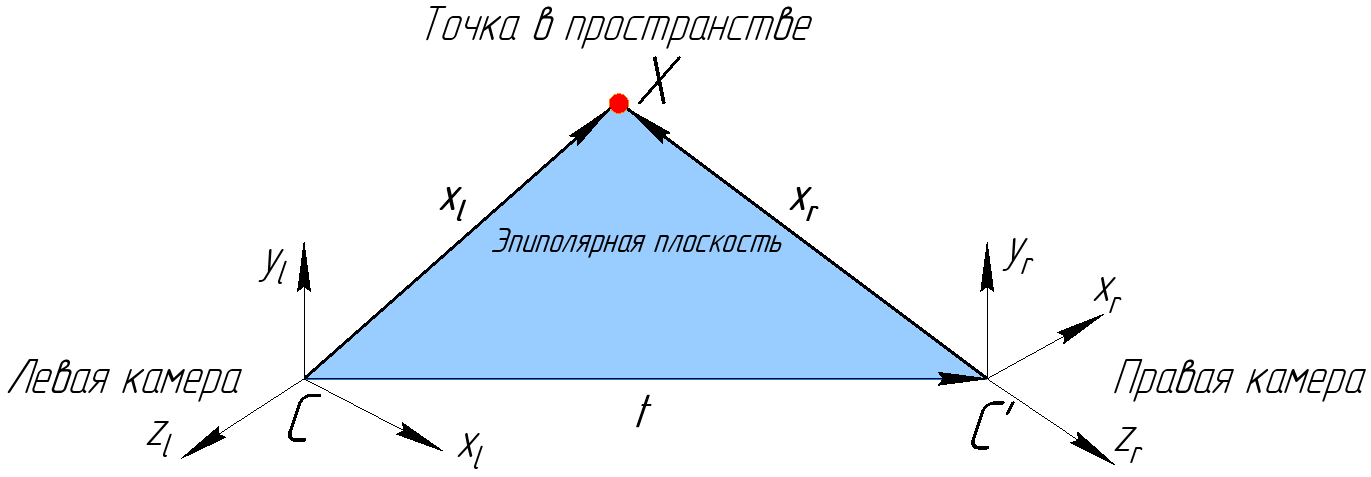


Рисунок 3.3 – Схема, задающая ограничения эпиполярной геометрии

Вышеописанные вектора имеют координаты, соответствующие координатам точки пространства в системах координат левой и правой камер. Вектор задает смещение центра одной камеры относительно другой. Тогда вектор нормали к эпиполярной плоскости будет задаваться следующим уравнением (3.7):

Отсюда определяется ограничение эпиполярной геометрии как скалярное произведение двух ортогональных векторов, а именно (3.8):

Представляя его в матричной форме, получим уравнение (3.9):

Переход из системы координат правой камеры к системе координат левой камеры для соответствующих друг другу точек задается уравнением (3.10):

где – элемент матрицы поворота.

Подставляя (3.10) в (3.9), получим уравнения (3.11), (3.12) и (3.13):

где – существенная матрица, представляющая собой произведение матриц поворота и перемещения.

Зная существенную матрицу, можно при помощи сингулярного разложения получить матрицы поворота и перемещения. Однако, в случае рассматриваемой системы, изначально известны исключительно координаты углов шахматной доски на изображении, поэтому, исходя из определенных ранее внутренних параметров камеры, необходимо перейти к однородным координатам точек на изображении, которые задаются уравнениями (3.14), (3.15):

где – координаты точки на изображении,

– матрица внутренних параметров левой камеры.

Тогда уравнение (3.12) примет следующий вид (3.16):

Откуда, учитывая, что и , получим уравнение (3.17):

где – фундаментальная матрица.

Определение фундаментальной матрицы производится при помощи 8-точечного алгоритма [49], после чего существенная матрица определяется при помощи уравнения (3.18), откуда разложением матрицы находятся внешние параметры камер.

3.3 **Триангуляция, нахождение трехмерных координат маркеров**

По данным, полученным в процессе калибровки системы камер, путем триангуляции определяются трехмерные координаты маркеров.

Учитывая вышеприведенные уравнения, задающие соотношения между системами координат двух камер, а также их плоскостей изображений, можно получить следующие уравнения (3.19), (3.20):

где - координаты точки на изображении.

Или, что то же самое, уравнения (3.21) и (3.22):

Откуда получим уравнения (3.23):

Обозначим матрицу как , а вектор справа от знака равенства как , получим (3.24):

Из уравнения (3.24) методом наименьших квадратов определяются трехмерные координаты точек согласно формуле (3.25):

3.4 **Выводы по данной главе**

В данном разделе исследуется алгоритм получения трехмерных координат ключевых точек по изображениям, полученным от пары камер.

Проведен анализ основных этапов алгоритма, математически описан каждый этап нахождения пространственных координат, Описанный алгоритм состоит из следующих пунктов:

– калибровка камер, получение их внутренних параметров,

– стереокалибровка пары камер, получение матриц, задающих преобразование из мировой системы координат в систему координат камеры,

– триангуляция, получение трехмерных координат ключевых точек.

4 **Программная реализация алгоритма**

Программная реализация алгоритма осуществлялась с использованием функционала библиотеки компьютерного зрения OpenCV (Open Computer Vision) поэтапно согласно представленной в предыдущей главе последовательности пунктов. Используются фидуциальные маркеры ArUco (Augmented reality University of Cordoba), структура которых облегчает их распознавание на изображении, что повышает скорость и качество их обнаружения. Библиотека для работы с маркерами ArUco включена в список библиотек, подключаемых к OpenCV. Для захвата движения применяется шестнадцать маркеров, закрепленных в следующих точках: голова, шея, правое и левое плечо, правый и левый локоть, правая и левая кисть, грудь, правое и левое бедро, тазобедренный сустав, правое и левое колено и правая и левая лодыжка. Выбор расположения маркеров объясняется не только необходимостью отметить ключевые точки тела, но и для упрощения дальнейшего представления полученных данных в качестве модели скелета. Схема расположения маркеров представлена на рисунке 4.1.

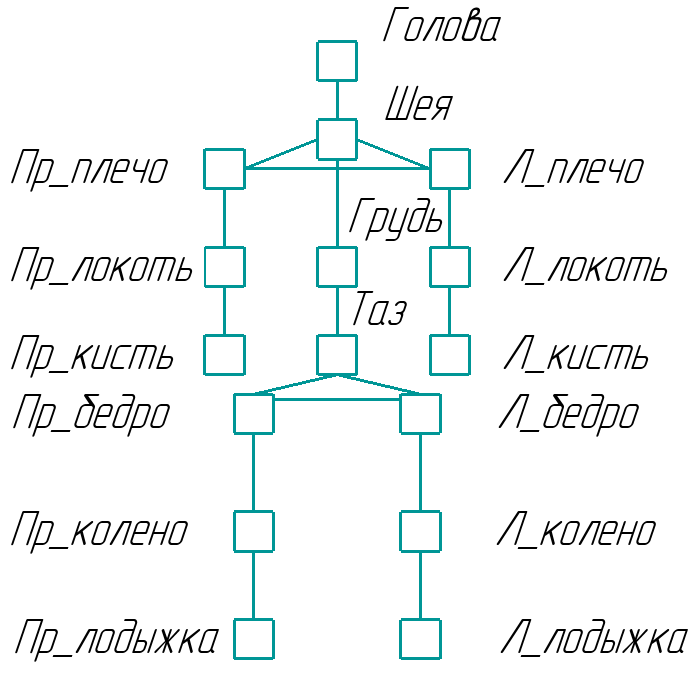


Рисунок 4.1 – Расположение маркеров на теле

4.1 **Калибровка камер**

Калибровка камеры проводилась при помощи шахматной доски согласно методу, изложенному Zhengyou Zhang[50]. Блок-схема алгоритма, по которому проводилась калибровка камер и коррекция дисторсии, представлена на рисунке 4.2.

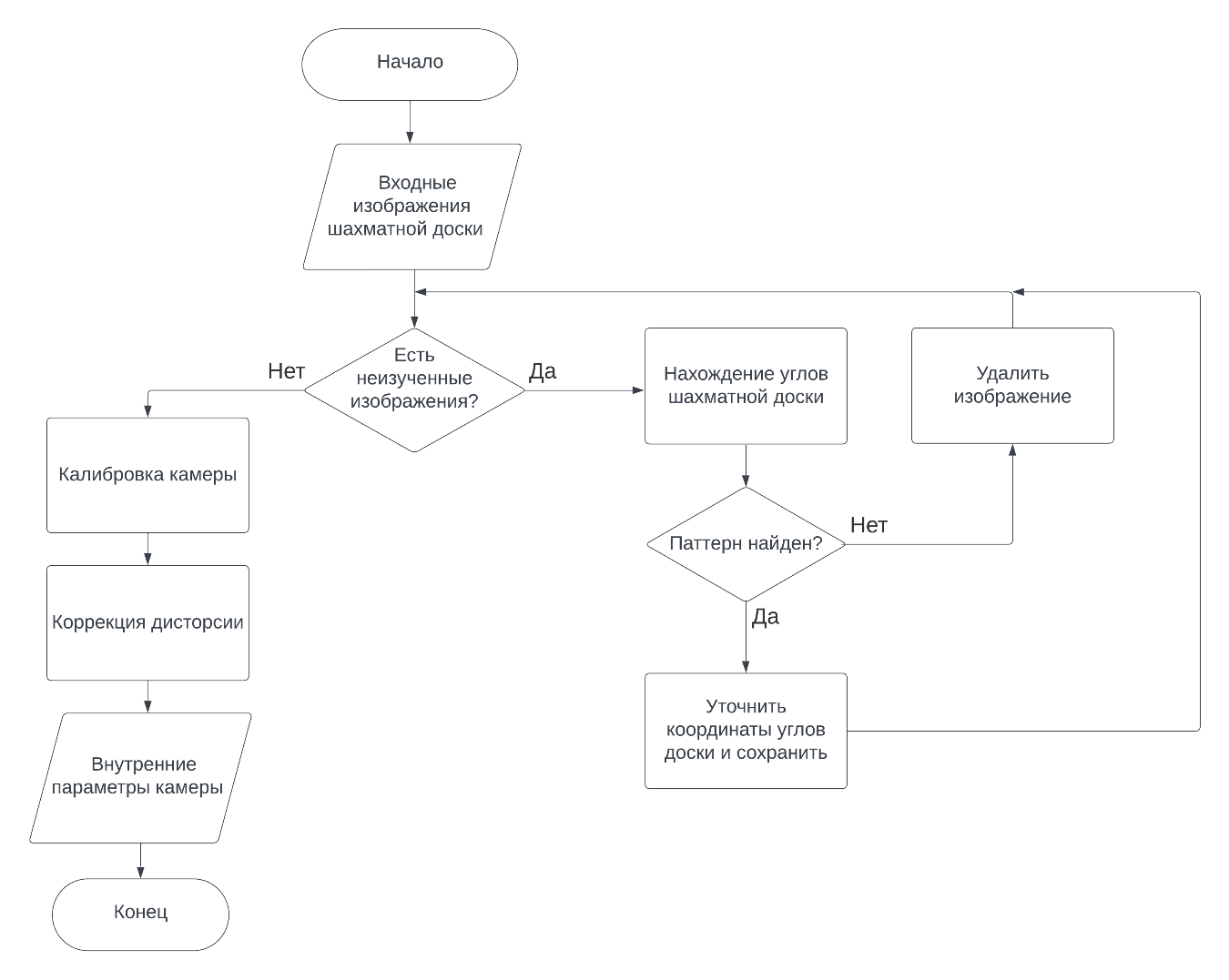


Рисунок 4.2 – Блок-схема алгоритма

На вход алгоритма подается выборка изображений с камеры, затем изображения переводятся в оттенки серого, и на каждом изображении при помощи встроенной функции cv::findChessboardCorners находятся углы шахматной доски. Затем, если паттерн был найден, производится уточнение координат углов с точностью до фрагментов пикселя с использованием функции cv::cornerSubPix, работа которой основана на алгоритме, представленном в [51], после чего координаты углов сохраняются в соответствующем контейнере. После того как были получены координаты углов шахматной доски на всех изображениях из выборки данные передаются для калибровки камеры в функцию cv::calibrateCameraRO, работающей согласно алгоритму, описанному в [52]. Использование данной функции вместо стандартной cv::calibrateCamera объясняется тем, что она позволяет значительно повысить точность определения внутренних параметров камеры при возможных неточностях в изготовлении и измерении калибровочного паттерна, а также в случаях неидеально плоских паттернов. Функция возвращает матрицу внутренних параметров камеры и вектор коэффициентов дисторсии, который используется затем в функции cv::undistort для коррекции радиальной и тангенциальной дисторсий. Процесс нахождения точек калибровочного паттерна представлен на рисунке 4.3.

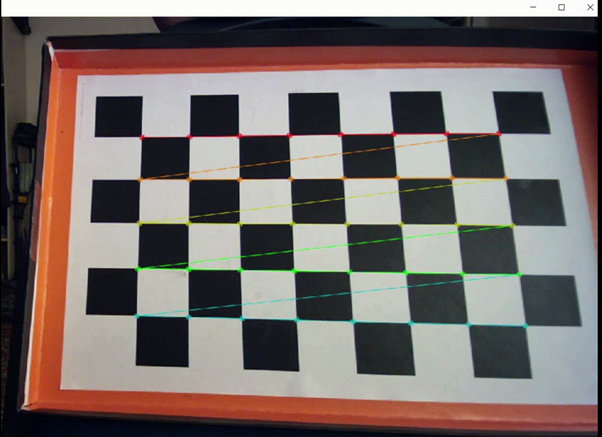


Рисунок 4.3 – Процесс детекции углов шахматной доски

Основным показателем, позволяющим отслеживать качество проведенной калибровки является ошибка репроекции, которая отмечает расстояние в пикселях между точкой калибровочного паттерна, обнаруженной на изображении и соответствующей точкой в мировой системе координат, спроецированной на изображение с использованием ранее полученных параметров калибровки. Результаты калибровки: матрица внутренних параметров, вектор коэффициентов дисторсии и итоговая ошибка репроекции – представлены на рисунке 4.4.

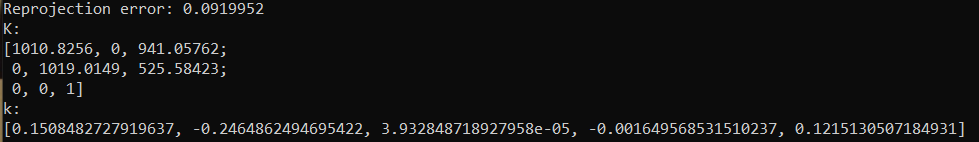


Рисунок 4.4 – Результаты калибровки

Для выборки из 50 изображений ошибка репроекции составила 0,092 для одной камеры и 0,12 для второй. Выборка формировалась путем многократного проведения калибровочного процесса с целью выявления изображений с наибольшей ошибкой репроекции, которые затем исключались из рассмотрения для снижения средней ошибки. По результатам калибровки производится коррекция дисторсии изображений, полученных с камеры, с использованием полученных коэффициентов.

4.2 **Калибровка пары камер**

Блок-схема алгоритма, согласно которому проводилась стереокалибровка, представлена на рисунке 4.5.



Рисунок 4.5 – Блок-схема алгоритма

Процесс калибровки пары камер осуществлялся при помощи функции cv::stereoCalibrate, которая принимает на вход координаты калибровочного паттерна в его системе координат, а также координаты углов шахматной доски на изображениях с двух камер. Данная функция вычисляет матрицы поворота и перемещения, задающие положение одной камеры относительно другой и возвращает значения фундаментальной и существенной матриц. Функция может осуществлять и полную калибровку камер, но данный подход не используется в силу большой размерности пространства параметров и большей уязвимости к шумам во входных данных согласно [53], поэтому внутренние параметры камер находятся отдельно. В случае разрабатываемой системы начало мировой системы координат принимается совпадающим с центром одной из камер, следовательно, матрицы камер будут иметь следующий вид (4.1), (4.2):

где – матрица камеры,

– матрица внутренних параметров камеры,

– матрица поворота системы координат второй камеры относительно системы координат первой камеры,

– вектор перемещения системы координат второй камеры относительно системы координат первой камеры.

Получение окончательных матриц камер производится конкатенацией матриц поворота и перемещения, в OpenCV это реализуется за счет функции cv::hconcat, осуществляющей горизонтальную конкатенацию, и их последующим умножением на матрицу внутренних параметров камеры. Результаты стереокалибровки, включая ошибку репроекции для той же выборки в пятьдесят изображений, для системы камер, в которой одна камера смещена относительно другой на расстояние 20 см исключительно по горизонтали представлены на рисунке 4.6 (данные приведены в миллиметрах).

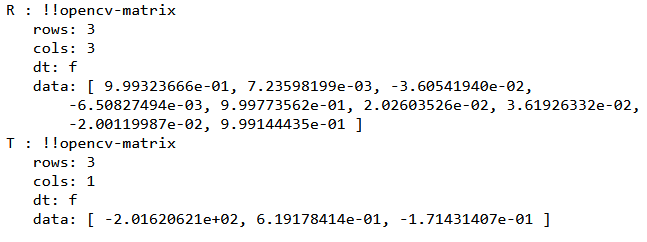


Рисунок 4.6 – Результаты калибровки пары камер

4.3 **Нахождение трехмерных координат маркеров**

Определение координат маркеров в пространстве осуществлялось при помощи триангуляции. Блок-схема алгоритма для одной пары кадров представлена на рисунке 4.7.

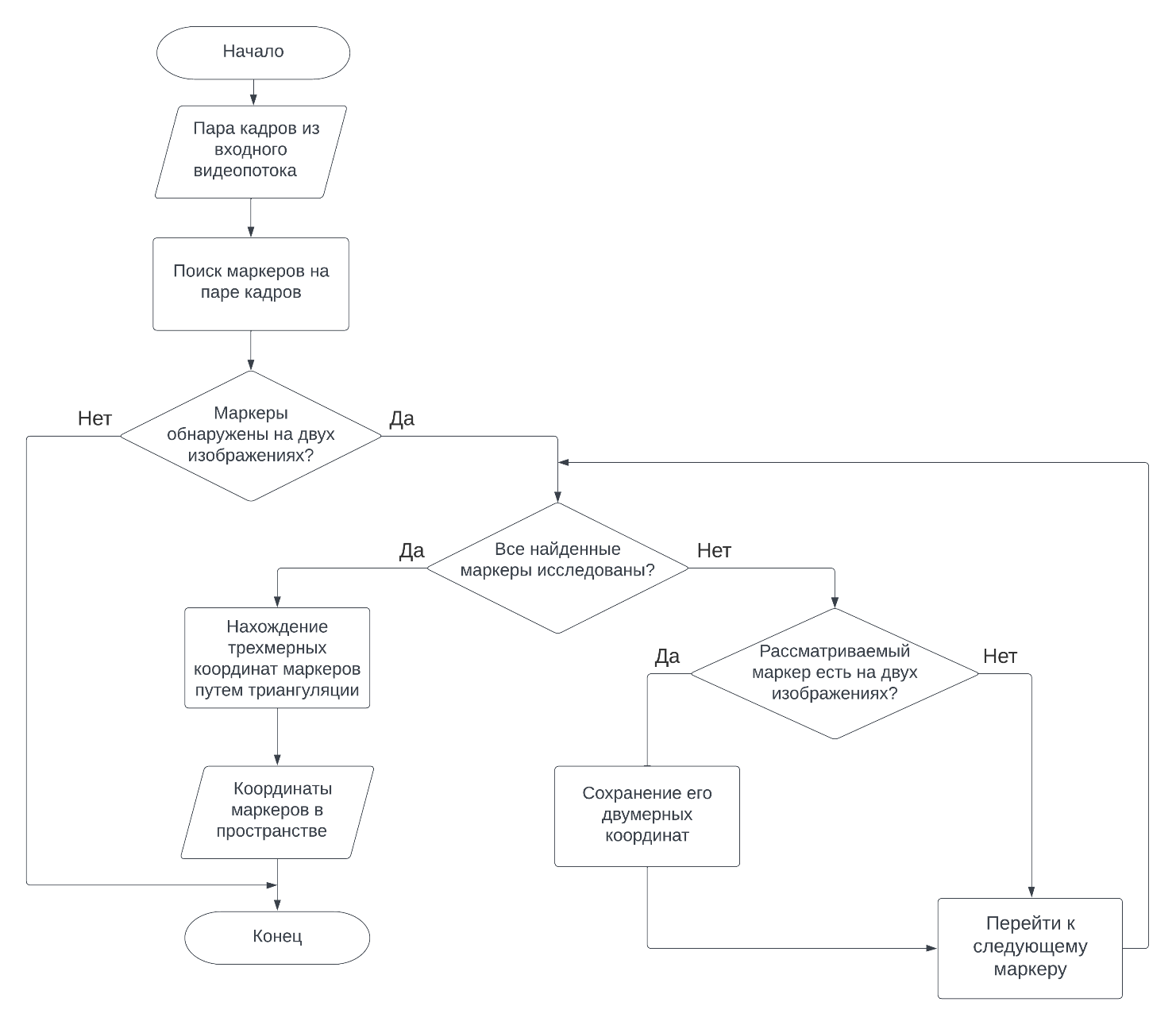


Рисунок 4.7 – Блок-схема алгоритма для пары кадров

Нахождение двумерных координат маркеров на изображениях с двух камер осуществлялось при помощи функции cv::aruco::detectMarkers, функция возвращает координаты углов маркеров на изображении, а также уникальный номер для каждого маркера. Если маркеры были успешно обнаружены на обоих кадрах, они выделяются при помощи функции cv::aruco::drawDetectedMarkers, а затем при помощи собственной функции equateVectors производится удаление из рассмотрения тех маркеров, что присутствуют лишь на одном из изображений, после чего по известным координатам углов находится центр каждого маркера. Затем матрицы двух камер, полученные в предыдущем пункте, а также двумерные координаты оставшихся маркеров передаются для проведения триангуляции в функцию cv::triangulatePoints. Данная функция принимает на вход координаты особых точек на двух изображениях, в рассматриваемом случае, ими служат координаты маркеров с одним уникальным номером. В результате функция выдает трехмерные координаты переданных ей точек, которые соответствуют координатам маркеров. В ситуации, представленной на рисунке 4.8, производится нахождение координат трех ключевых точек, полученные данные выводятся на изображение.



Рисунок 4.8 – Результат работы программы для трех ключевых точек

4.4 **Выводы по разделу**

В данном разделе приведена программная реализация алгоритма захвата фидуциальных маркеров и получения их трехмерных координат, описанного в предыдущем разделе, и его основных этапов. Приведены блок-схемы основных этапов, приведена схема закрепления маркеров в ключевых точках тела. Алгоритм реализован при помощи библиотеки OpenCV.

ЗАКЛЮЧЕНИЕ

В ходе данной работы был проведен анализ систем захвата движений на основе камер, выявлены их достоинства и недостатки, в результате чего в качества базиса разрабатываемой системы были выбраны системы на основе нательных маркеров в силу большей точности нахождения координат ключевых точек.

Далее была разработана система захвата движений на основе веб-камер и фидуциальных маркеров ArUco, описана основная концепция системы и ключевые этапы разработки, включающие в себя калибровку камер и нахождение трехмерных координат маркеров путем триангуляции. В ходе дальнейшей работы планируется повысить надежность детекции маркеров, так как система уязвима к изменениям освещения и частичному перекрытию маркеров, сформировать модель скелета на основе получаемых от системы данных для визуального представления результатов, а также провести сравнение разработанной системы с аналогами.

Рисуно

Рис

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 Lanzoni, D. [et. al], Markerless Motion Capture and Virtual Reality for Real-Time Ergonomic Analysis of Operators in Workstations with Collaborative Robots: a preliminary study. // In: Advances on Mechanics, Design Engineering and Manufacturing IV – 2023.

2 Racinskis, P. [et al.], A Motion Capture and Imitation Learning Based Approach to Robot Control, J. Appl. Sci. – 2022 – Vol. 12.

3 Xu, W. [et al.], Natural teaching for humanoid robot via human-in-the-loop scene-motion cross-modal perception", J. Industrial Robot – 2019 – Vol. 46 – P. 404-414.

4 Agethen, P [et al.], Using marker-less motion capture systems for walk path analysis in paced assembly flow lines, J. Procedia CIRP – 2016 – Vol. 54 – P. 152–157

5 Akhavian, R., Behzadan, A.H., Productivity analysis of construction worker activities using smartphone sensors. // In Proceedings of the 16th International Conference on Computing in Civil and Building Engineering, Osaka, Japan – 2016.

6 Bortolini, M. [et al.], Motion Analysis System (MAS) for production and ergonomics assessment in the manufacturing processes, J. Comput. Ind. Eng: A. - 2020 – Vol. 139

7 Ferrari, E. [et al.], A. Motion Analysis System for the digitalization and assessment of manual manufacturing and assembly processes, J. IFAC-PapersOnLine: A. – 2018 – Vol. 51 – P. 411–416

8 Bohannon, R.W. [et al.], Reliability and validity of pendulum test measures of spasticity obtained with the Polhemus tracking system from patients with chronic stroke. J. Neuroeng. Rehabil.: A. – 2009 – Vol. 6

9 Brents, C. [et al.], Low Back Biomechanics of Keg Handling Using Inertial Measurement Units. In Software Engineering in Intelligent Systems – Springer Science and Business Media: Berlin/Heidelberg, Germany, 2018 – P. 71–81.

10 Needham, L. [et al.], The accuracy of several pose estimation methods for 3D joint centre localisation., J. Sci Rep: A. – 2021 – Vol. 11

11 Scimmi, L.S. [et al.], Implementing a Vision-Based Collision Avoidance Algorithm on a UR3 Robot. // In Proceedings of the 2019 23rd International Conference on Mechatronics Technology (ICMT), Salerno, Italy, – 2019 – P. 1–6

12 Plantard, P. [et al.], Validation of an ergonomic assessment method using Kinect data in real workplace conditions, J. Appl. Ergon.: A. – 2017 – Vol. 65 – P. 562–569

13 Park, Y. [et al.], Development of a wearable sensing glove for measuring the motion of fingers using linear potentiometers and flexible wires, J. IEEE Trans. Ind. Inform.: A. – 2014 – Vol. 11 – P. 198–206

14 Cardoso, M. [et al.], Biomechanical investigation of prolonged driving in an ergonomically designed truck seat prototype, J. Ergonomics: A. – 2017 – Vol. 61 – P. 367–380

15 Yueting Zhuang [et al.], A modern approach to intelligent animation - Zhejiang University Press, Hangzhou, 2009 – P.69 – 87

16 Jha, A. [et al.], Imitation learning inindustrial robots. // In Proceedings of the Advances in Robotics on-AIR ’17, New Delhi, India – 2017 – P. 1–6.

17 Kun-Do Lee, Hyung-Soon Park. Real-Time Motion Analysis System Using Low-Cost Web Cameras and Wearable Skin Markers, J. Frontiers in Bioengineering and Biotechnology: А. – 2022 – Vol. 9.

18 Khan MH [et al.], Marker-Based Movement Analysis of Human Body Parts in Therapeutic Procedure, J. Sensors: A. – 2020 – Vol. 20 – P. 3312

19 Pfister A. [et al.], Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis, J. Journal of Medical Engineering & Technology: A. – 2014 – Vol. 38 – P. 274-280

20 Otte, K. [et al.], Accuracy and reliability of the kinect version 2 for clinical measurement of motor function, J. PLoS ONE: A. – 2016 – Vol. 11

21 Brennan, A. [et al.] Quantification of Inertial Sensor-Based 3D Joint Angle Measurement Accuracy Using an Instrumented Gimbal., J. Gait & Posture. – 2011 – Vol. 34 – P. 320–323.

22 Filippeschi, A. [et al.] Survey of Motion Tracking Methods Based on Inertial Sensors: A Focus on Upper Limb Human Motion. J. Sensors. – 2017 – Vol. 17 – P. 1257.

23 Yu-Ichi Ohta [et al.], Color information for region segmentation, J. Computer Graphics and Image Processing: A. – 1980 – Vol. 13 – P. 222 – 241

24 Canny J., A Computational Approach to Edge Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence: A. – 1986 – Vol. 8 – P. 679 – 698

25 Nobuyasu N. [et al.], Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras, J. Frontiers in Sports and Active Living, A. – 2020 – Vol. 2.

26 Needham L. [et al.], The accuracy of several pose estimation methods for 3D joint centre localisation., J. Sci Rep: A. – 2021 – Vol. 1

27 Bhrigu K. [et al.], Development and evaluation of a new methodology for Soft Tissue Artifact compensation in the lower limb, Journal of Biomechanics: A. – 2021 – Vol. 122.

28 Camomilla V. [et al.], Human movement analysis: The soft tissue artefact issue, Journal of Biomechanics: A. – 2017 – Vol. 62 – P. 1 – 4.

29 Gløersen, Øyvind, Federolf, Peter, Predicting Missing Marker Trajectories in Human Motion Data Using Marker Intercorrelations, J. PloS one: A. – 2016 – Vol. 11.

30 Chiari, L [et al], Human movement analysis using stereophotogrammetry. part 2: Instrumental errors, J. Gait and Posture: A. – 2005 – Vol. 21 – P. 197 – 211

31 Zhe Cao [et al.] Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields – 2017

32 Zhe Cao [et al.] OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields – 2019

33 S.-E. Wei [et al.], Convolutional pose machines // CVPR 2016 – 2016

34 Seethapathi N. [et al.], Movement science needs different pose tracking algorithms, ArXiv, abs/1907.10226 – 2019

35 Colyer S. [et al.] A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System, J. Sports Med: A. – 2018 – Vol. 24.

36 Khoshelham K. [et al.], Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications, J. Sensors: A. – 2012 – Vol. 12 – P. 1437 – 1454

37 Lange R., Seitz P., Solid-state time-of-flight range camera, J. IEEE Journal of Quantum Electronics – 2001 – Vol. 37 – P. 390-397

38 Giancola, S. [et al.], A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies – Springer: Berlin, Germany, 2018

39 Shotton J. [et al.], Real-time human pose recognition in parts from single depth images, CVPR – 2011 – P. 1297-1304

40 Neupane, C. [et al.], Evaluation of Depth Cameras for Use in Fruit Localization and Sizing: Finding a Successor to Kinect v2, J. Agronomy: A. – 2021 – Vol. 11 – P. 1780

41 Halmetschlager-Funek G. [et al.], An Empirical Evaluation of Ten Depth Cameras: Bias, Precision, Lateral Noise, Different Lighting Conditions and Materials, and Multiple Sensor Setups in Indoor Environments, J. IEEE Robotics & Automation Magazine: A. – 2019 – Vol. 26 – P. 67-77

42 Vit A, Shani G., Comparing RGB-D Sensors for Close Range Outdoor Agricultural Phenotyping, J. Sensors: A. – 2018 – Vol. 18 – P. 4413.

43 Mejia-Trujillo J. D. [et al.], Kinect™ and Intel RealSense™ D435 comparison: a preliminary study for motion analysis, // IEEE International Conference on E-health Networking, Application & Services – 2019 – P. 1-4

44 Jianwei Niu [et al.] A Novel Method of Human Joint Prediction in an Occlusion Scene by Using Low-Cost Motion Capture Technique, J. Sensors – 2020 – Vol. 20.

45 Edwards M., Green R., Low-Latency Filtering of Kinect Skeleton Data for Video Game Control, The 29th International Conference on Image and Vision Computing New Zealand – 2014 – P. 190 – 195.

46 Jody Shu [et al.], Application of extended Kalman filter for improving the accuracy and smoothness of Kinect skeleton-joint estimates, Journal of Engineering Mathematics – 2014 – Vol. 88 – P. 161 – 175.

47 Zennaro S., Evaluation of Microsoft Kinect 360 and Microsoft Kinect One for robotics and computer vision applications, Master’s thesis, University of Padova, Italy – 2014.

48 Brown, D. C., Decentering distortion of lenses, J. Photogrammetric Engineering – 1966 – Vol. 32 – P. 444–462.

49 Hartley R. I., "In defense of the eight-point algorithm," // IEEE Transactions on Pattern Analysis and Machine Intelligence – 1997 – Vol. 19 – P. 580-593.

50 Zhang Z., A flexible new technique for camera calibration // IEEE Transactions on Pattern Analysis and Machine Intelligence – 2000 – Vol. 22 – P. 1330-1334.

51 Forstner W., A fast operator for detection and precise location of distincs points, corners and center of circular features // In Proc. of the Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland – 1987 – P. 281–305.

52 Strobl K. Hirzinger, H. G., More accurate pinhole camera calibration with imperfect planar target // IEEE International Conference on Computer Vision Workshops, Barcelona, Spain – 2011 – P. 1068-1075.

53 Bradski, G., The OpenCV Library, J. Dr. Dobb's Journal of Software Tools – 2000.