**Разработка системы захвата движений человека для телеуправления гуманоидным роботом**

Введение

Тренды мировой роботизации указывают на повышающееся значение взаимодействия человека и робота в рамках рабочего и жизненного пространств. Часто наиболее приспособленными к работе в условиях повседневной среды обитания являются человекоподобные роботы в силу схожести их морфологии с человеческой. Они используются в телемедицине для дистанционного проведения операций ([1], [2]), помощи в реабилитации и проведения лечения при отсутствии возможности личного присутствия врача ([3]). Высокая вариативность человеческих характеров, которые определяют их реакции и действия в процессе взаимодействия с роботом, значительно усложняют процесс программирования гуманоидных роботов, так как лишает возможности заранее учесть развитие всех возможных вариантов событий. Телеуправление гуманоидными роботами представляется решением данной проблемы, а возможность его построения на основе захвата движений предоставляет оператору удобный и естественный интерфейс управления.

Большинство используемых технологий управления гуманоидными роботами включают три основных этапа: (1) Захват движений человеческого тела, получение необходимых координат суставов (2) Обработка полученных данных, (3) Применение обработанных данных для управления гуманоидным роботом. Данная базовая, упрощенная схема представлена на рисунке 1.

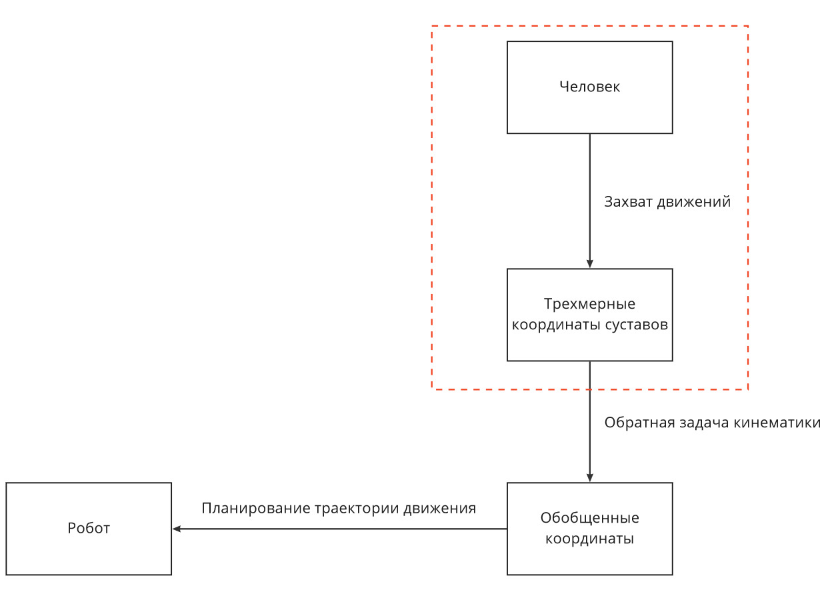


Рисунок 1 – Схема управления, основанная на захвате движений

В данной работе рассматривается первый этап – захват движений человека и определение трехмерных координат суставов. Наиболее совершенными системами считаются оптические системы 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.

Одними из часто применяющихся систем с меньшей стоимостью являются системы на основе инерциальных датчиков – акселерометров и гироскопов, положения частей тела в которых определяются путем интегрирования полученных данных. Но интегрирование шумов в данных датчиках приводит к постепенному снижению точности определения координат и углов поворота суставов ([4], [5]). Так, согласно [4] максимальный дрейф угла в инерциальном измерительном блоке доступной цены, закрепленном на кардановом подвесе с тремя степенями свободы, составлял , что значительно искажало измерения ориентации суставов, поэтому в данной работе решено было остановиться на рассмотрении систем, использующих камеры для захвата движений.

Методы без использования маркеров, применяемые для телеуправления роботами, зачастую базируются на использовании Kinect ([6], [7], [8], [9]), состоящего из RGB-камеры, а также инфракрасных излучателя и приемника, которые позволяют определять расстояние до объекта в каждой точке. Система широко применяется в силу своей экономической доступности, компактности и простоты использования, однако данные, полученные при помощи данной камеры, подвержены значительному фазовому дрожанию [10], и на точность определения положений суставов существенное влияние оказывают шумы, а также возможное перекрытие или временное исчезновение частей тела из кадра ([11], [12]), к тому же первая версия системы непригодна для использования вне помещений в условиях высокой освещенности (Kinect for robotics). В Kinect One устранено большинство проблем предшественника, но его стоимость увеличилась в несколько раз, что затрудняет его повсеместное использование, поэтому видится рациональным разработать алгоритм захвата движений, сопоставимый по точности измерений с системами Kinect, но предлагающий более дешевое и доступное решение.

Целью работы является разработка системы захвата движений на основе маркеров и веб-камер массового сегмента.

Конечная практическая цель с точки зрения использования – разработка системы захвата движений, сравнимой по качеству позиционирования с системами Kinect, учитывающего их недостатки, но обладающей существенно меньшей стоимостью и большей гибкостью.

Маркеры выбраны потому, что они позволяют сильно упростить задачу поиска и определения положения объекта, при этом их можно разместить на человеке, используя повязки или нанести на производственную одежду.

(не поставляется в связи с санкциями, что свидетельствует о том, что продукция считается важной для научно-технического развития)

(стоимость, не всегда возможно повесить датчики)

1. Обзор существующих методов

1.1. Захват движений с использованием маркеров

Данный метод предполагает распознавание на изображении, полученном с камеры, нательных маркеров, которые отличаются между собой по цвету [2], нанесенному на них паттерну [1] или же представляют собой маркеры дополненной реальности [3].

1.1.1. Цветные маркеры

Результатом данного метода является получение координат центров маркеров на изображении. Очень важным становится выбор правильного цветового пространства для последующей оценки цветовых диаграмм и выделения контуров. Согласно [4] выбирается наиболее эффективная система из трех ортогональных признаков, определенных по формулам 1.1 – 1.3.

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 1 представлен результат предсказаний положения нагрудного маркера.

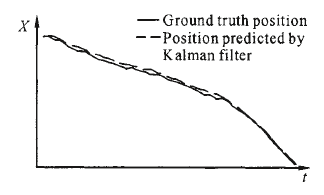
**

Рисунок 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. 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 представлено сравнение традиционного и улучшенного алгоритмов.

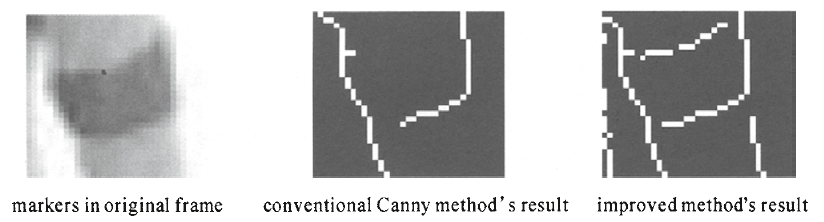


Рисунок 2 – Сравнение методов Кэнни

Границы на полученных бинарных изображениях не являются непрерывными, следовательно, не формируют необходимых замкнутых контуров. Для решения данной проблемы применяется преобразование Хафа, (Почему-то не удалось найти нигде такого описания данного преобразования, которое используется здесь: «Hough transform converts the image to a parameter space, where the parameters are represented by grids, and the best parameter grid is searched for». Особенно непонятен момент с сетками, вместо них везде используется аккумулятор Хафа, про который я и стал писать далее. Также опустил описание круглого маркера, так как преобладают все-таки прямоугольные маркеры.), 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 следующим образом:

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

– угол между нормалью и осью абсцисс.

(Здесь опять же в статье написано про сетки и оптимизацию метода при помощи регулирования их толщины.)

Результат применения преобразования Хафа представлен на рисунке 3.

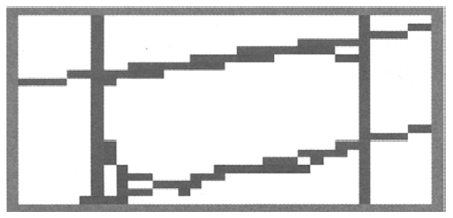


Рисунок 3 – Преобразование Хафа

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, (В статье написано: «After Hough transform, eight lines with highest probabilities are reserved», но непонятно, какие вероятности имеются в виду, поэтому я рассматривал их как значения 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:

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:

(В статье указано другое условие: , здесь используется нормальное уравнение прямой, и, если условие 1.6. абсолютно понятно, то второе условие, кажется, совсем не работает для параллельности, так как, если прямые параллельны, углы и при вычитании 180, два условия должны быть одновременно меньше , для которого тоже нет объяснения, поэтому я ввел ее как некую минимальную величину близкую к нулю)

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

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

Условие перпендикулярности:

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 прямой .

Так как и симметричны относительно :

Следовательно, путем сложения и получится:

(Непонятно, почему при сложении и превратились в и . Предполагаю, что, так как все рассматривается относительно предсказанного центра , в первых двух уравнениях тоже должны быть и )

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 :

(Не указано, что такое S, возможно, это какая-то целевая функция, тоже никак не объясняется, указано лишь, что она убывает, весовые коэффициенты просто приведены без указания каких-либо возможных значений)

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. (Проверку прямоугольников по гистограммам опускаю как побочный метод)

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 следующим образом:

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. (Здесь я пропускаю детальную процедуру калибровки)

В качестве начальных точек используется треугольник «грудь-левое плечо-правое плечо», так как данные точки меньше всего подвержены перекрыванию и деформации. (Далее в статье приведен анализ графика, на котором в качестве графика с окружностями назван график «Z coordinate», но в самом тексте написано: «the circle curve shows the speed centroid of the starting triangle along the Z direction». К тому же чуть выше на той странице написано, что результаты довольно посредственны и не отражают реального движения человека, но, когда далее дается пояснения к следующему графику (у меня это рисунок 4), написано, что метод работает хорошо)

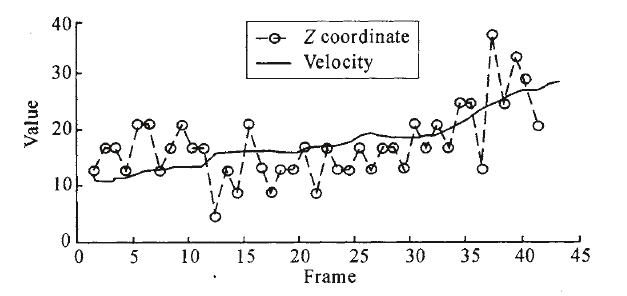
Если координаты точек стартового треугольника на *n*-ом кадре равны и скорость каждой точки равна , следовательно интервал поиска для кадра может быть определен как . Учитывая, что изменение координат треугольника происходит постепенно, можно ограничить интервал значений координат и следующим образом:

где .

Помимо этого, важным критерием расчета является тот факт, что длины всех ребер стартового треугольника должны быть постоянными (). Также учитывается непрерывность движения, количественно определяемая через расчет ускорения (). отвечает за разницу направлений векторов нормали в двух последовательных кадрах. Исходя из этого, оптимальное положение может быть определено следующим образом, при условии :

(Здесь не поясняется, что такое , но можно предположить, что это какая-то оптимальная координата, при этом коэффициенты не пояснены совсем, дано лишь их соотношение)

На рисунке 4 представлены значения координаты Z, полученные с использованием данного метода. Недостатком является тот факт, что в методе используются параметры, значения которых задаются вручную и могут оказывать значительное влияние на результат. Для выставления параметров, например, при ходьбе, когда движение с постоянной скоростью вдоль оси Z может быть представлено в виде прямой линии, параметр приравнивается к нулю, а параметру задается значение большее единицы.



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 5, таких точек может быть две.

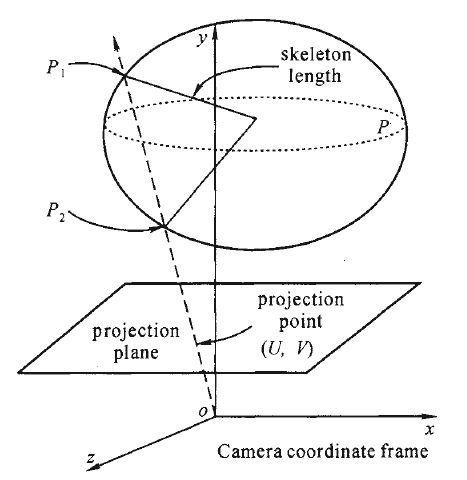


Рисунок 5 – Неоднозначность полученных результатов

(В статье есть цитата: «and finally uses an object function based on global motion smoothing assumption to select a single pose.». Так и не удалось найти, что же означает объектная функция и motion smoothing assumption, поэтому решил пропустить их упоминание здесь)

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 6 приведены значения скоростей суставов руки для тридцатитрехкадрового клипа, на котором запечатлен момент прыжка.

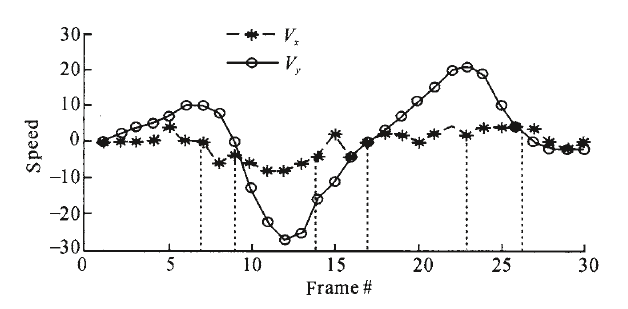


Рисунок 6 – График движения руки

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 :

(в статье она задана как «smoothing criterion function», возможно, она отвечает за сглаживание или плавность движений.)

где – угловое ускорение на трех кадрах.

На основе предыдущего уравнения получено:

определяет степень сглаживания(плавности?) в пределах кадров. Таким образом, для каждого кадра строится граф, представленный на рисунке 6, включающий в себя начальную точку и конечную точку . Таким образом, поиск удовлетворяющих вершин преобразуется в решение проблемы поиска кратчайшего пути в графе.

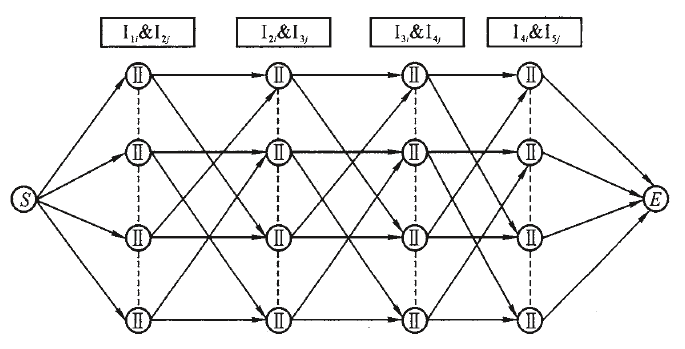


Рисунок 6 – Граф точек

1.1.2. Маркеры с нанесенным паттерном

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 1.

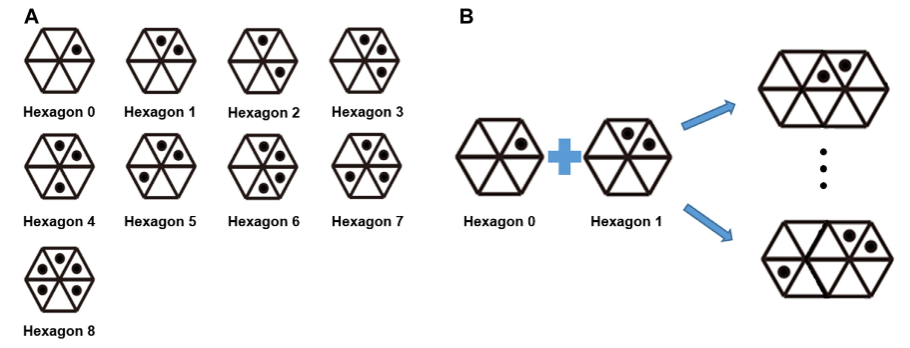


Рисунок 1 – Используемые шестиугольники

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.

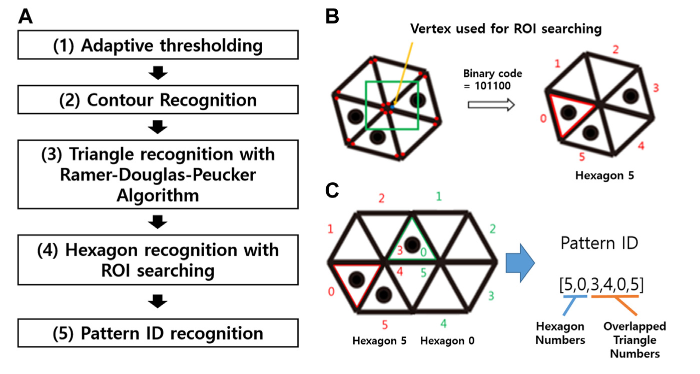


Рисунок 2 – Распознавание маркеров

Процесс оценки положения тела человека состоит из двух этапов: анализа статичного положения тела и анализа тела в движении. На первом шаге, представленном на рисунке 3, 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.

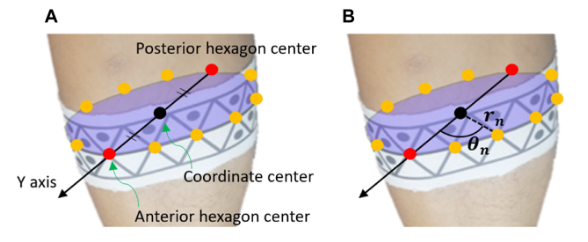


Рисунок 3 – Определение направления осей

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.

Преимущества:

Большая по сравнению с системами без маркеров точность [15], [16];

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

3D-координаты могут быть получены с использованием только одной камеры [17];

Недостатки:

Артефакты, вызванные движением мягких тканей, отделяющих маркер от кости, известные также как STA (Soft Tissue Artifacts) [18, 19];

Снижение точности измерений из-за возможного перекрытия маркеров [20, 21]:

Наличие маркеров может привести к снижению естественности движений оператора;

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

2. Захват движений без использования маркеров

2.1. Библиотека OpenPose

Данная система согласно [5] принимает на вход изображение размером и в качестве результата выдает 2D координаты ключевых точек тела каждого человека, изображенного на фотографии. Схема работы метода представлена на рисунке 2.

Метод основан на создании карт достоверности (*Confidence maps*) и полей сходства частей тела (*Part Affinity Fields*).

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 1.

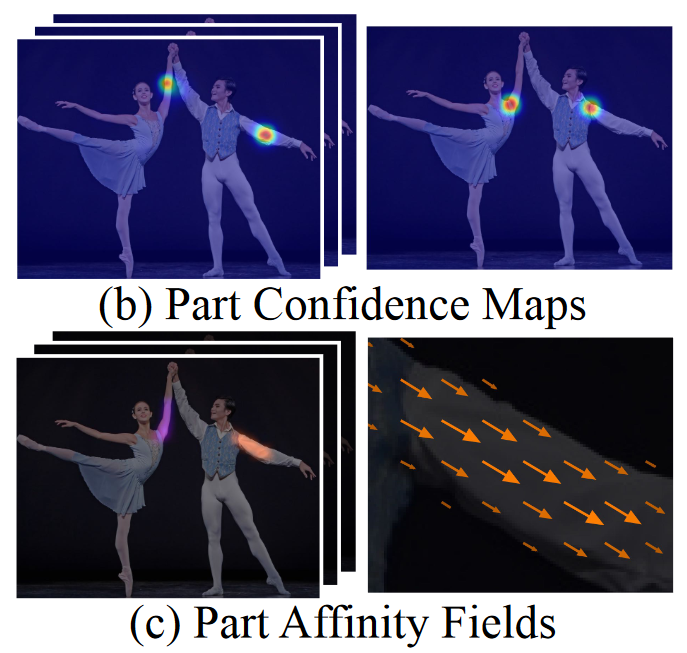


Рисунок 1 – Карта достоверности и поле сходства

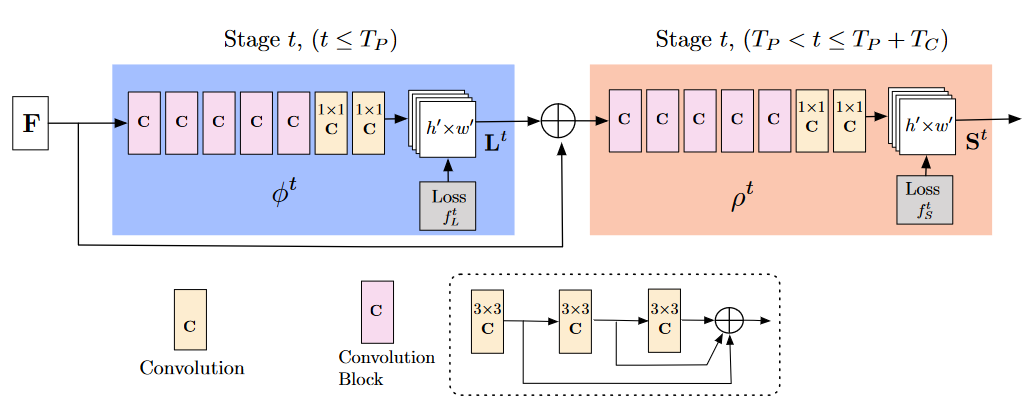


Рисунок 2 – Структура сверточной нейронной сети

Для получения карты признаков, являющейся входными данными данной архитектуры, используются первые десять слоев сверточной нейронной сети VGG-19. На данном этапе нейронная сеть создает набор полей сходства частей тела, , 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:

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:

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 выглядят следующим образом:

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 [4], увеличивающий значение градиента после каждой стадии. Итоговая функция потерь:

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:

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 [22];

Данные методы реализуются с использованием двух и более камер, что повышает их стоимость и требует калибровки;

Точность данных методов уступает системам, основанным на маркерах [15], [16], [17];

Наибольшие ошибки в измерениях наблюдаются вследствие неправильной разметки эталонных данных на тренировочных датасетах [22], [23].

2.2 Система Kinect

Kinect состоит из RGB-камеры, а также инфракрасных излучателя и приемника. Получение расстояний до объекта производится методом триангуляции согласно [14]. 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 1 отражает соотношение между расстоянием от камеры до точки объекта , расстоянием до эталонной плоскости и полученным смещением .

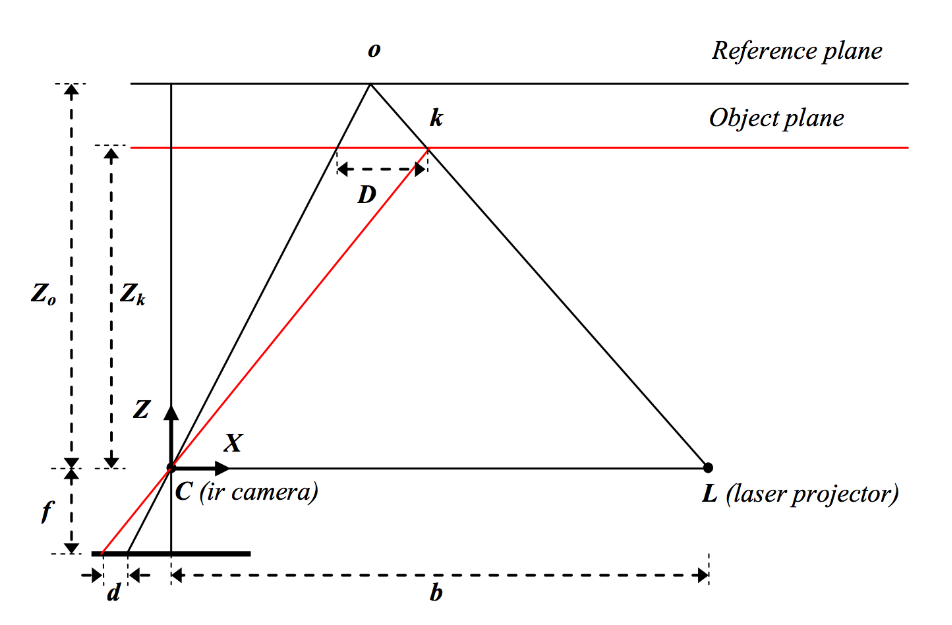


Рисунок 1 – Схема получения расстояния до объекта

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:

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 является базовым для получения расстояния до объекта по известному значению диспарантности , при этом параметры и определяются в процессе калибровки. Координаты и могут быть получены следующим образом:

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 [15], [16]. Излучатель испускает определенную модулированную волну. Полученная приемником волна после отражения имеет некоторый сдвиг по фазе . Допустим, что излучается сигнал:

где – частота модуляции.

Следовательно, на приемнике будет сигнал:

где – постоянный сдвиг;

– амплитуда.

Демодуляция сигнала производится путем взаимной корреляции данных сигналов. Корреляционная функция задается уравнением:

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

Следовательно, сдвиг по фазе определяется как:

Следовательно, расстояние до объекта равно:

где – скорость света.

Для распознавания положения тела полученное с камеры глубины изображение человека разбивается на цветные сегменты, представленные на рисунке 2, 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 – 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 вычисляется следующим образом:

где – глубина в данном пикселе на изображении ;

– параметры, описывающие смещения.

На рисунке 3 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.

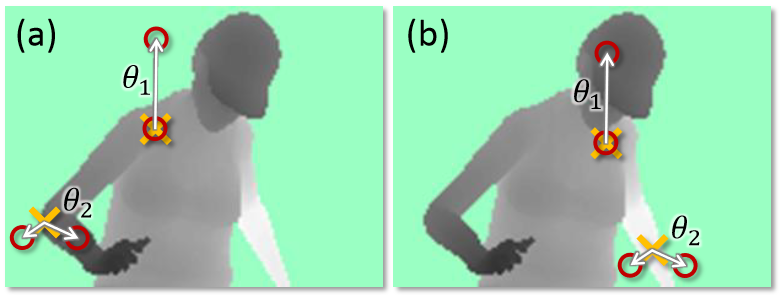


Рисунок 3 – Расчет признаков

Для классификации используется метод случайного леса, представляющий собой ансамбль деревьев принятия решений. На рисунке 4 представлен лес, состоящий из деревьев, каждое из которых включает в себя узловые и краевые вершины. Каждая узловая вершина состоит из признака и порогового значения .

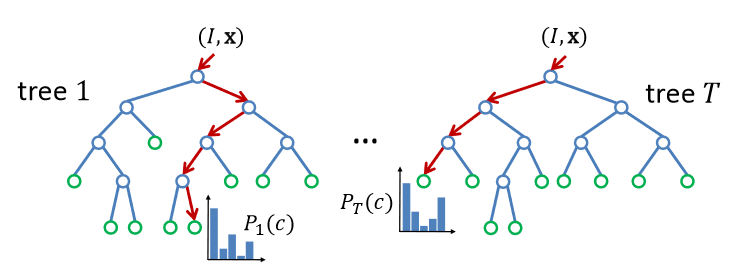


Рисунок 4 – Метод случайного леса

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:

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:

где – положение в мировых координатах;

– количество пикселей в изображении;

– вес пикселя;

– проекция пикселя изображения в мировых координатах;

– пропускная способность.

Вес каждого пикселя учитывает распределение для каждого пикселя и площадь поверхности, заключенной в нем:

Преимущества:

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 [10];

Снижение точности измерений при перекрытии конечностей или выходе из поля зрения [11], [12];

Освещенность снимаемой сцены значительно влияет на точность полученных данных, первое поколение системы способно захватывать вне помещений только предметы, находящиеся в тени [13];

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

Таблица 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 [15, 16, 22, 23];  Последовательные временные интервалы рассматриваются как статистически независимые, что приводит к ошибкам оценки положения тела в последовательных кадрах [22];  Малое количество систем в открытом доступе  Для получения 3D-координат требуется две и более камеры, что повышает стоимость и требует дополнительной калибровки |
| Камеры глубины(Kinect и др.) | Одного устройства достаточно для получения координат  Не ограничивает движений оператора  Простота использования системы | Высокая уязвимость к шумам [11]  Снижение точности измерений при перекрытии конечностей или выходе из поля зрения [11, 12];  Значительное снижение точности в зависимости от освещенности сцены [10, 13]  Данные подвержены значительному фазовому дрожанию [10];  Значительное снижение точности вне помещений  Высокая стоимость, прекращены поставки |
| На основе маркеров | Большая по сравнению с предыдущими системами точность [15, 17]  Могут быть реализованы на основе камер с низкой стоимостью  3D-координаты могут быть получены с использованием только одной камеры [17] | Снижение точности измерений в случае перекрытия маркеров [20, 21]  Возникновение артефактов, вызванных движением мягких тканей, отделяющих маркер от кости (Soft Tissue Artifacts) [18, 19]  Более длительная настройка системы, расположение маркеров в нужных местах  Возможное снижение естественности движений оператора |

Заключение

В данной работе были рассмотрены существующие оптические системы, применяемы для захвата движений человека, проанализированы их достоинства и недостатки, представленные в таблице 1.

Системы на основе сверточных нейронных сетей представляют наиболее естественное решение с точки зрения оператора, так как не ограничивают его движений, и представляют потенциально наиболее широко доступные для массового пользователя решения, так как не требуют специализированных камер и дополнительных приспособлений для получения координат суставов, однако точность данных систем все еще уступает точности тех, что основаны на маркерах. Большинство используемых алгоритмов не учитывают межкадровую корреляцию, рассматривая последовательные промежутки времени как статистически независимые, что приводит к неправильной оценке позы человека на следующих друг за другом кадрах, дополнительные ошибки в измерениях возникают в связи с неправильной разметкой эталонных данных.

Методы, основанные на использовании камер глубины, таких как Kinect, широко применяются в силу удобства и простоты их использования, так как одно устройство позволяет получить необходимую информацию о координатах, но уязвимость к шумам и зависимость от освещенности снимаемых помещений, а также высокая цена и прекращение поставок ограничивают возможность использования данных систем.

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. Системы представляют экономически выгодное решение, так как могут быть реализованы на основе наиболее доступных камер, а разработка специальных маркеров позволяет получать 3D-координаты суставов с использованием только одной камеры.

Список использованных источников

1 Mayer H. [et al.] “A system for robotic heart surgery that learns to tie knots using recurrent neural networks,” J. Advanced Robotics – 2008 – Vol. 22 – P. 1521-1537.

2 Song C. [et al.] “RoMAN-MD: 6 DOF humanoid arm for medical applications,” In Advanced Robotics and its Social Impacts, IEEE Workshop – 26 October 2010 – P. 123-123.

3 Csala E. [et al.] “Application of the NAO humanoid robot in the treatment of marrow-transplanted children,” Cognitive Infocommunications, IEEE 3rd International Conference – 2012 – P. 655-659.

4 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.

5 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.

6 Michieletto S. [et al.] Teaching humanoid robotics by means of human teleoperation through RGB-D sensors, J. Robotics and Autonomous Systems – 2016 – Vol. 75 – P. 671-678.

7 Yang Xu [et al.] Robot teaching by teleoperation based on visual interaction and extreme learning machine, J. Neurocomputing – 2018 – Vol. 275 – P. 2093-2103.

8 G. Peng [et al.], Teleoperation control of Baxter robot based on human motion capture, IEEE International Conference on Information and Automation - 2016 – P. 1026-1031.

9 R Dilip [et al.], Real-time Motion Control of Robot Using Kinect, J. Phys.: Conf. Ser. – 2021.

10 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.

11 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.

12 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.

13 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.

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

15 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.

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

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 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.

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

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

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

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

23 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. Sports Med - Open 4, 24 (2018