Общество с ограниченной ответственностью «Цифровые решения»

УДК: 621.397

Регистрационный №АААА-А19-119042390070-0

Инв. №01-2019

УТВЕРЖДАЮ

Генеральный директор

Матвеев Д.В.

" " 201\_г.

**М.П.**

ОТЧЕТ

о выполнении НИОКР по теме:

"Разработка обучающей базы изображений и видеопоследовательностей с учетом специфики спортивного зала. Разработка тестовой базы изображений и видеопоследовательностей с учетом специфики спортивного зала. Разработка алгоритмов детектирования объектов на видеоданных с диагонального ракурса камеры. Разработка алгоритмов трекинга объектов на видеоданных с диагонального ракурса камеры."

(договор 2962ГС1/45326 от 01.04.2019)

(промежуточный)

|  |  |
| --- | --- |
| Научный руководитель работ | Хрящев В.В.  подпись, дата |

Ярославль, 2019

1. **Реферат**

Отчет 47 стр., 14 илл., 0 табл., 0 приложений, 30 источников

детектирование объектов, сопровождение объектов, машинное обучение, компьютерное зрение, нейронные сети

Цель работы на данном этапе – сбор базы и изображений, а также разработка и тестирование алгоритмов детектирования и сопровождения объектов на основе алгоритмов машинного обучения, составляющих основу системы видеоаналитики спортивных мероприятий. Основными результатами работы являются:

* для обучения и тестирования алгоритмов были подготовлены базы изображений баскетболистов, состоящие из, соответственно, 1500 и 500 изображений с разрешением от 50x100 пикселей до 200x400 пикселей;
* изображения из базы были подвергнуты геометрическим искажениям с целью увеличения объема данных для обучения нейронных сетей;
* разработан алгоритм детектирования изображений баскетболистов с диагонального ракурса видеокамеры;
* разработан алгоритм сопровождения изображений баскетболистов с диагонального ракурса видеокамеры;
* проведено первичное тестирование разработанных алгоритмов на одном из спортивных объектов города Ярославля.

Областями применения разработанных алгоритмов являются комплексы спортивной видеоаналитики, которые будут использованы спортивными клубами, многофункциональными спортивными комплексами.

Работы по данному этапу НИОКР были выполнены в полном объеме в соответствии с техническим заданием и календарным планом. Полученные результаты будут использованы на следующих этапах НИОКР и в коммерческой деятельности ООО «Цифровые решения».

**СПИСОК ИСПОЛНИТЕЛЕЙ**

|  |  |  |
| --- | --- | --- |
| **Должность** | **Подпись** | **ФИО, номер раздела** |
| Генеральный директор, ООО «Цифровые решения», к.т.н. |  | Матвеев Д.В. (подразделы 3.1 – 3.4) |
| Научный руководитель работ, к.т.н. |  | Хрящев В.В. (подразделы 3.3 – 3.4) |
| Программист ООО «Цифровые решения» |  | Казина Е.М. (подраздел 3.3) |
| Консультант ООО «Цифровые решения» |  | Федькина А.А. (подразделы 3.1-3.2) |

**СОДЕРЖАНИЕ**

|  |  |
| --- | --- |
| **1. Реферат** | 2 |
| **2. Введение** | 5 |
| **3. Основная часть** | 9 |
| 3.1. Разработка обучающей базы изображений и видеопоследовательностей с учетом специфики спортивного зала | 9 |
| 3.2. Разработка тестовой базы изображений и видеопоследовательностей с учетом специфики спортивного зала | 13 |
| 3.3. Разработка алгоритмов детектирования объектов на видеоданных с диагонального ракурса камеры | 14 |
| 3.3.1 Обнаружение объектов | 14 |
| 3.3.2 Создание модели игрока | 16 |
| 3.3.3 Модель обучения | 17 |
| 3.3.4 Удаление ложноположительных обнаружений | 27 |
| 3.3.5 Удаление зала | 27 |
| 3.3.6 Удаление судьи | 29 |
| 3.3.7 Разделение игроков на команды | 30 |
| 3.3.8 Определение позиции игрока | 33 |
| 3.4. Разработка алгоритмов трекинга объектов на видеоданных с диагонального ракурса камеры | 36 |
| 3.4.1 Модель наблюдения | 36 |
| 3.4.2 Стратегия отслеживания | 37 |
| 3.4.3 Отслеживание объекта как последовательная байесовская оценка | 37 |
| 3.4.4 Отслеживание объектов как двойной режим двухстороннего байесовского вывода | 39 |
| **4. Заключение** | 43 |
| **5. Список использованных источников** | 45 |

**2. ВВЕДЕНИЕ**

В компьютерном зрении одной из основных проблем является автоматическое детектирование и классификация объектов [1-5]. В спортивной видеоаналитике одной из основных проблем является объективный анализ производительности игрока. Хотя физические возможности каждого игрока могут быть предварительно проверены специальными исследованиями, производительность конкретного игрока, а также всей команды можно оценить лишь во время матча. Правильное использование статистики и интеллектуальный анализ данных относительно редко встречается в этой области в настоящее время. Основным препятствием на пути внедрения методов интеллектуального анализа данных в спорте является неприятие новых технологий, встречающееся в большинстве спортивных организаций. В данной НИОКР мы предлагаем новый подход к обнаружению игроков и определению их позиций на игровой площадке в соответствии с заранее определенными областями на примере баскетбола. Определение позиции игрока является очень важным параметром в анализе прошедших игр, а также в подготовке к новым.

Обнаружение игрока – довольно сложная проблема. Основная причина состоит в том, что баскетбол – очень динамичный вид спорта, в котором игроки постоянно меняют свое положение и позы. Цель данного этапа НИОКР предложить надежную и эффективную систему, которая обеспечит обнаружение игрока и определение его положение во время игры посредством видеоаналитики. Одной из основных проблем на пути создания подобных систем является ситуация перекрытия контура человека другими объектами [6-10].

Обнаружение игрока будем осуществлять с помощью методов, которые разбивают внешний вид объектов на шаблоны локальных деталей вместе с геометрическими ограничениями на расположение деталей [11]. У игроков этот алгоритм позволяет обнаружить руки, ноги, голову и туловище игрока. Данные части тела будут использоваться в качестве важных параметров для дальнейшего анализа.

Обучение модели выполняется с использованием изображений игроков, характеризующихся четырнадцатью точками (лодыжки, колени, бедра, запястья, локти, плечи, подбородок и кожа головы). Область вокруг каждой точки представляет собой образец определенной категории объектов [12]. По каждой области мы вычисляем гистограмму ориентированных градиентов (HOG), чтобы создать вектор признаков [13]. Причиной введения таких векторов признаков является тот факт, что они менее восприимчивы к изменениям цвета из-за различных условий освещения. Это также является одной из причин, почему наш подход может быть более устойчив, чем подходы, основанные на использовании цвета [13].

Векторы признаков для каждой категории являются входными данными для алгоритма машины опорных векторов (SVM), который создает их модель. В процессе обнаружения полное изображение преобразуется в вектор признаков с использованием алгоритма HOG [13-14]. Чтобы обнаружить эти объекты, мы применяем обученную модель для каждой категории по всем позициям и масштабам на наблюдаемом изображении. Этот тип детектора можно рассматривать как классификатор, который принимает положение на изображении и его масштаб в качестве входных данных. Такой детектор определяет, есть ли образец требуемой категории в наблюдаемых положении и масштабе. На данном этапе НИОКР обнаружение игрока в баскетбольной игре происходит при видеозаписи в формате FullHD (который сейчас является стандартом в трансляции спортивных событий).

После применения фильтров к новому изображению мы обнаруживаем баскетболистов. Рядом с игроками также существует определенное количество ложных срабатываний. В основном это люди из зала и судьи. Чтобы исключить людей из зала, будем полагать, что они не находятся на игровой площадке. Чтобы удалить зал, мы исключаем те обнаруженные фигуры, чьи лодыжки не находятся в границах баскетбольного поля. Обнаружение лодыжек обеспечивается моделью, основанной на частях тела и применяемой в этом исследовании.

Внутри площадки, среди игроков, мы также выявляем судей. В нашем исследовании они представляют ложноположительные обнаружения. Поскольку судьи во всех играх лиги носят форму одного цвета, мы можем рассчитать их цветовую гистограмму и сравнить ее с цветовыми гистограммами других обнаруженных объектов. Если гистограммы похожи, такой объект представляет рефери, и он исключается из дальнейшего рассмотрения, а оставшиеся объекты являются игроками. Мы объединяем их в две отдельные группы, каждая из которых представляет одну команду. Кластеризация основана на значении насыщенности цвета, которое рассчитывается с использованием цветового пространства HSV [11-15].

Завершающим этапом является определение расположения на площадке обнаруженных игроков. По этой причине необходимо дополнительное обнаружение линий на поле. Помимо горизонтальных и вертикальных границ площадки, мы также детектируем границы области под баскетбольной корзиной (выделяются цветом). Таким образом, мы можем знать точное местоположение пяти точек на нашем изображении (четыре точки в каждом углу краски и точка, созданная пересечением горизонтальной и вертикальной границ площадки). Определяя эти точки и принимая во внимание фактический размер поля, можно рассчитать пространственное преобразование, которое отображает точки с изображения на реальные точки поля. С помощью этого преобразования мы можем рассчитать положение каждого игрока. Чтобы нанести на карту всех игроков, необходимо определить их точное местоположение на изображении. Мы достигаем этой цели с использованием точки между ногами игроков, которая и определяет его местоположение.

Для построения эффективных алгоритмов спортивной видеоаналитики на основе методов машинного обучения на данном этапе НИОКР решаются следующие основные задачи:

* разработка обучающей базы изображений и видеопоследовательностей с учетом специфики спортивного зала.
* разработка тестовой базы изображений и видеопоследовательностей с учетом специфики спортивного зала.
* разработка алгоритмов детектирования объектов на видеоданных с диагонального ракурса камеры.
* разработка алгоритмов трекинга объектов на видеоданных с диагонального ракурса камеры.

Полученные результаты будут использованы на следующих этапах НИОКР и в коммерческой деятельности ООО «Цифровые решения».

**3. ОСНОВНАЯ ЧАСТЬ**

**3.1. Разработка обучающей базы изображений и видеопоследовательностей с учетом специфики спортивного зала.**

Для разработки системы спортивной видеоаналитики на первом этапе работ была приобретена антивандальная купольная камера GV-EVD3100 - разработанная в соответствии со стандартами IK10+ и IP67, что позволяет применять данную купольную камеру для подобных систем, без риска выведения из строя системы при попадании спортивными снарядами, такими как баскетбольный мяч. Камера оснащена автоматическим ИК-фильтром и ИК-светодиодами, а также светочувствительной 1/2.8” CMOS-матрицей с прогрессивной разверткой, способной обеспечить цветное изображение в почти полной темноте. Регуляция положения объектива по 3 осям позволяет размещать данную камеры как на потолке спортивного зала, так и на стене, и получать ракурсы съемки, необходимые для работы систем спортивной видеоаналитики. Внешний вид камеры приведен на рисунке 1.

|  |
| --- |
| Image result for GV-EVD3100 |
| Рисунок 1 – Внешний вид камеры GV-EVD3100 |

Высоту подвеса камеры на стене спортивного зала ЯрГУ им. П.Г. Демидова можно оценить с помощью фотографий точки наблюдения, приведенных на рисунке 2.

|  |  |
| --- | --- |
| C:\Users\Wicher\Pictures\IMG-3b7c5bf2607ffb5682ee06ab5af3199d-V_.jpg | C:\Users\Wicher\Pictures\IMG-2471a49b4ce47b0187a489789d842756-V_.jpg |

Рисунок 2 – Монтаж камеры в спортивном зале ЯрГУ

Рассматриваемая камера обладает следующими техническими характеристиками:

* максимальное разрешение 2048x1536;
* минимальная освещенность для получения цветного изображения 0,01;
* скорость затвора от 1/5 до 1/8000 сек.;
* отношение сигнал/шум 54 дБ;
* поддержка WDR Pro;
* динамический диапазон до 100 дБ;
* вариофокальный мегапиксельный объектив с фокусным расстоянием 3-9 мм, максимальной апертурой F/1.7 и горизонтальным полем зрения 96-36 градусов;
* механическая диафрагма P-Iris с ограничением закрытия диафрагмы, которое предотвращает размытие видео в тех случаях, когда автоматическая диафрагма чрезмерно реагирует на яркий свет;
* поддержка двух одновременных видеопотоков с кодеками H.264 и MJPEG;
* максимальное число кадров в секунду: 30;
* возможные разрешения для главного видеопотока: 2048x1536, 1600x1200, 1280x960, 640x480, 320x240 – для соотношения сторон 4:3, 1920x1080, 1280x720, 640x360, 448x252 – для соотношения сторон 16:9, 1280x1024, 640x512, 320x256 для соотношения сторон 5:4;
* возможные разрешения для второстепенного видеопотока: 960x720, 640x480, 320x240 – для соотношения сторон 4:3, 640x360, 448x252 – для соотношения сторон 16:9, 640x512, 320x256 для соотношения сторон 5:4;
* поддержка PoE по стандарту IEEE 802.3af;
* поддержка протокола RTSP для передачи видеопотока.

Для обучения алгоритмов искусственного интеллекта была проведена съемка реальной игры баскетбольных команд в спортивном зале ЯрГУ им. П.Г. Демидова из которой была подготовлена база изображений, состоящая из 1500 изображений баскетболистов, с разрешением от 50x100 пикселей до 200x400 пикселей. Примеры изображений из обучающей базы приведены на рисунке 3.

|  |
| --- |
| C:\Users\Wicher\Pictures\dataset.jpg |
| Рисунок 3 – Примеры изображений из обучающей базы |
|  | |
| Данные изображения будут использованы, как для обучения традиционных методов машинного обучения (например SVM), так и в дальнейшем – для алгоритмов на основе сверточных нейронных сетей. Во втором случае планируется дальнейшее расширение обучающей базы по мере накопления видеоматериалов. | |

**3.2. Разработка тестовой базы изображений и видеопоследовательностей с учетом специфики спортивного зала.**

Разделение на обучающее и тестовое множество изображений осуществляется путем деления выборки в определенной пропорции, например обучающее множество - две трети данных и тестовое - одна треть данных. Этот способ используется для выборок с большим количеством примеров. Если же выборка имеет малые объемы, обычно применяют специальные методы, при использовании которых обучающая и тестовая выборки могут частично пересекаться

Для тестирования алгоритмов спортивной видеоаналитики была подготовлена тестовая база изображений баскетболистов, состоящая из 500 изображений, с разрешением от 50x100 пикселей до 200x400 пикселей. Примеры тестовой базы баскетболистов приведены на рисунке 4.

|  |
| --- |
| ../Pictures/test_set.jpg |
| Рисунок 4 – Примеры изображений из тестовой базы баскетболистов |

**3.3. Разработка алгоритмов детектирования объектов на видеоданных с диагонального ракурса камеры.**

**3.3.1 Обнаружение объектов**

Обнаружение объектов в основном достигается с использованием либо деформируемых моделей, либо моделей на основе частей. Типичная модель на основе частей – модель созвездия [24]. Части ограничены редким набором местоположений, а геометрические аргументы определяются распределением Гаусса. Напротив, модель графической структуры [23] определяет проблему соответствия, когда детали имеют индивидуальную стоимость соответствия в плотном наборе возможных расположений.

Модель, используемая на данном этапе НИОКР, по большей части основана на графических структурах [23, 28]. Они используют плотный набор возможных местоположений и масштабов на наблюдаемом изображении. Для каждого из этих мест и масштабов оценка определяется путем применения к ним набора фильтров. Геометрическая конфигурация определяется набором функций стоимости деформации, которые регулируют связь между частью тела и ее родителем. Расходы на деформацию часто визуализируются в виде пружин. Напротив, традиционные модели для распознавания объектов используют части, которые характеризуются только расположением пикселей [19]. Наш подход существенно отличается от некоторых других исследований, связанных с баскетболом [16, 20], которые используют цвет в качестве основы для обнаружения игрока.

Значительные изменения в форме и внешнем виде, которые происходят в результате экстремальных изменений точки зрения, не могут быть хорошо смоделированы с использованием двумерных деформируемых моделей [22]. Следовательно, в игру вступает использование смешанных моделей. В них для декодирования фронтальных и боковых видов лиц или некоторых других объектов используются несколько шаблонов.

Сопоставление деформируемой модели с изображением объекта является сложной задачей оптимизации. Методы локального поиска требуют инициализации, которая примерно равна правильному решению. Общий подход в моделях на основе частей заключается в ограничении положения частей лишь небольшим набором возможных мест [24]. Древовидные графические структуры [23] позволяют использовать динамическое программирование и обобщенные преобразования расстояний. С их использованием можно получить эффективный поиск по всем возможным конфигурациям на изображении. Эти методы сопоставления используются в наших исследованиях для определения оценки при некоторой конфигурации фильтра по наблюдаемому изображению. Древовидная модель, полученная из графической структуры, представлена ​​на рисунке 5.

|  |
| --- |
|  |
| Рисунок 5 – Иллюстрированная структура с пружинами и древовидной моделью |

Модели на основе частей параметризуются по внешнему виду каждой детали и геометрической модели, включающей пространственные отношения между частями. В полностью контролируемой системе все части обучающих изображений размечены. Это позволяет моделировать обучение, которое выполняется с помощью более простых методов [23]. Геометрическую модель в нашем исследовании изучают, применяя кластеризацию k-средних по местоположениям всех частей из обучающей базы изображений

Дискриминационные методы обучения определяют параметры модели, минимизирующие погрешность алгоритма обнаружения по набору обучающих образов. Подобные подходы напрямую оптимизируют границы принятия решений между положительными и отрицательными примерами. Это является одной из причин успеха простой модели, использующей дискриминационные методы, такие как детекторы Виолы-Джонса и Далала-Триггса [21]. Для этого мы используем SVM со скрытыми переменными.

**3.3.2 Создание модели игрока**

Первая цель нашей работы – обнаружение игроков в баскетбольном матче. Для этой цели мы использовали статические изображения из записанных сюжетов.

Оценка осанки всего тела чрезвычайно трудна из-за множества степеней свободы, которые должны быть оценены. Баскетбол, как очень динамичный вид спорта, еще больше усложняет оценку позы, поскольку конечности могут двигаться в самых разных направлениях. Они сильно различаются из-за различий в форме тела и изменений в точке зрения, которые могут привести к ложному появлению сокращения. Это приводит к большей сложности. В заключении, нам приходится искать изображения с большим количеством повернутых и расположенных в разных ракурсах шаблонов. Yang и Ramanan решили эту проблему, введя новое представление для моделирования семейства аффинно-деформированных шаблонов, которое называется смесью неориентированных графических структур [28]. Они дополняют классические «пружинные» модели ограничениями совместного вхождения, которые благоприятствуют определенной комбинации деталей. Такие ограничения могут моделировать локальную жесткость (ситуация, когда две части на одной конечности имеют одинаковую ориентацию).

**3.3.3 Модель обучения**

Обучение осуществляется по алгоритму, который основан на смеси частей. При таком подходе ориентация частей тела достигается с помощью ряда шаблонов, которые создаются для каждой части. Оптимальное количество шаблонов варьируется в зависимости от того, наблюдаем ли мы за головой, рукой, ногой и т.д. Это число определяется перекрестной проверкой, а когда обучение основано на одном наборе баскетболистов, оно имеет пять или шесть шаблонов на деталь. Это отличается от стандартного метода, в котором для любой части тела используется один шаблон [23]. Моделирование деталей ранее упомянутым способом обеспечивает гораздо большую гибкость.

Следующим шагом является преобразование точек в области, которые будут использоваться при обучении. Чтобы определить ширину области, необходимо определить расстояние между каждой точкой и точкой, представляющей ее родителя. Расстояние между двумя точками вычисляется с использованием расстояния Евклида:

|  |  |
| --- | --- |
| . |  |

Мы берем среднее значение расстояния, поэтому все области имеют одинаковый размер. Эти области показаны на рисунке 6.

|  |  |
| --- | --- |
| Описание: C:\Users\Wicher\Pictures\crop_14.png | Описание: C:\Users\Wicher\Pictures\crop_26.png |
| (а) 14 точек, используемых в обучении | (б) 26 точек, используемых в обучении |
| Описание: C:\Users\Wicher\Pictures\crop_26_big.png |  |
| (в) Области, используемые в обучении | (г) Точечные метки, используемые в обучении |
| Рисунок 6 – Модель тренировки на основе изображений баскетболиста | |

Через области обучения мы рассчитали векторы признаков HOG [13]. Основной причиной использования алгоритма HOG является необходимость создания надежного набора функций, позволяющего четко различать игрока или его часть, даже если фон или освещение не являются постоянными. Основная идея этого алгоритма состоит в том, что части объекта и их форму можно хорошо описать, используя распределение локальных градиентов интенсивности. Реализация этого алгоритма достигается путем разделения наблюдаемого изображения на небольшие пространственные области, называемые ячейками. Для каждой ячейки определяется локальная одномерная гистограмма направлений градиента или ориентации краев по пикселям в данной ячейке. Комбинации значений гистограммы создают новое представление объекта. Чтобы добиться более низкой чувствительности к изменениям освещения, теней и т.д., мы применили нормализацию оттенков. Это было сделано путем измерения накопления «энергии» локальных гистограмм в больших пространственных областях, называемых блоками. Накопленная энергия используется для нормализации всех ячеек в блоке.

*Скрытые переменные.* В наборе тренировочных образов части находятся в суставах. Это означает, что у нас есть метки для позиций частей , но у нас нет меток для типов деталей . Типами деталей могут быть, например, ориентации деталей (горизонтальные или вертикальные). Чтобы получить метки типа в сочлененной модели, необходимо вручную определить структуру ребер . Эта структура будет связывать местоположения парных частей по их близости (рисунок 6г). Можно предположить, что типы деталей соответствуют разному относительному расположению некоторой детали относительно ее родителя в . Например, боковые ориентированные руки появляются рядом с локтями, в то время как ориентированные вниз руки появляются ниже локтей. Это означает, что мы можем использовать относительное местоположение как дополнительную подсказку, которая может помочь нам идентифицировать типы, которые представляют ориентации в этой статье. Предположим, что -й тренировочный образ имеет помеченные местоположения суставов . Пусть представляет относительное положение части относительно ее родителя на наблюдаемом изображении. Для каждой части мы выполняем кластеризацию k-средних по ее относительному расположению в обучающем наборе , чтобы получить кластеров. Каждый кластер соответствует образцам частей с согласованными относительными местоположениями и, следовательно, с согласованной ориентацией [28]. Метки типов деталей определяются в соответствии с их принадлежностью к кластеру.

Поскольку типы частей выводятся эвристически, можно рассматривать как скрытую переменную. Это скрытая проблема SVM, и цель состоит в том, чтобы найти максимальную гиперплоскость, которая будет отделять любую точку в одном из двух классов (класс точек, который представляет наблюдаемую часть, и класс точек, которые не представляют эту часть). Гиперплоскость должна создаваться таким образом, чтобы примеры отдельных классов были разделены максимально возможным зазором. Гиперплоскость определяется параметрами , которые представляют вектор, нормальный к разделяющей гиперплоскости, и параметром , который представляет смещение гиперплоскости от начала координат вдоль вектора нормали . Это показано на рисунке 7.

Ширина зазора составляет (это можно определить геометрически), поэтому целью является минимизация . Скрытая задача SVM может быть эффективно решена с помощью метода спуска координат [16]. Оценочная функция является линейной по параметрам модели .

Если мы напишем , то функцию оценки можно записать как произведение точек между параметрами модели, и вектор признаков, где вектор признаков представляет собой гистограмму ориентированных градиентов:

|  |  |
| --- | --- |
|  |  |

В нашем исследовании мы изучаем модель следующего вида:

|  |  |
| --- | --- |
|  | (1) |

где:

В предыдущем уравнении представляет собой постоянное значение и является слабой переменной. Слабые переменные введены для того, чтобы допустить неправильно маркированные примеры. Это примеры, которые неправильно классифицируются по гиперплоскости разделения. Теперь оптимизация становится компромиссом между большим запасом и небольшой ошибкой.

|  |
| --- |
|  |
| Рисунок 7 – Графическое представление алгоритма SVM |

Обученные модели применяются к дискретным картам характеристик, чтобы вывести правильную конфигурацию деталей путем сравнения их оценок. Карта объектов – это массив, элементы которого представляют собой векторы пространственных объектов , рассчитанные по плотной сети местоположений на наблюдаемом изображении. Это реализуется путем разделения окна изображения на небольшие пространственные области, которые называются «ячейками». Для каждой ячейки локальная 1-D гистограмма направлений градиента рассчитывается по пикселям ячейки. Лучшие результаты достигаются путем применения нормализации контраста. Это можно сделать, накапливая меру энергии локальной гистограммы в больших пространственных областях, которые называются «блоками», и используя результаты для нормализации всех ячеек в блоке. Нормализованные блоки дескрипторов называются дескрипторами гистограммы ориентированного градиента (HOG) [21]. Чтобы создать карту объектов для произвольного изображения, мы рассчитываем дескрипторы HOG по всему изображению.

Наша цель – определение оценки модели по всем позициям и масштабам на изображении. Оценка является точечным произведением фильтра и подокна карты объектов. Это может быть достигнуто с помощью пирамиды объектов, которая определяет карту объектов для конечного числа масштабов в фиксированном диапазоне. Элементная пирамида может быть создана путем построения стандартной пирамиды изображений с помощью повторного сглаживания и подвыборки. Пирамида изображения показана в верхней части рисунка 8. Следующий шаг – вычисление карты объектов на каждом уровне пирамиды изображения. Это показано в нижней части рисунка 8.

Выборка масштаба в пирамиде объектов определяется параметром , который определяет количество уровней в октаве. Другими словами, – это количество уровней, на которое нам нужно спуститься в пирамиде, чтобы получить карту объектов, рассчитанную с удвоенным разрешением другой. Точная выборка масштабного пространства важна для получения высокой производительности [22]. Все части тела созданы для представления клеток HOG в размере 5×5. Наблюдая за обучающими изображениями, мы пришли к выводу, что голова представляет часть тела с наименьшими изменениями в размерах. Это связано с тем, что угол наклона головы относительно камеры практически постоянен (в отличие от рук или ног). Поэтому мы создали гистограмму размера головы в пикселях для всех игроков, которые вошли в процесс обучения и тестирования. Изображения игроков получаются из видеороликов, транслируемых через телевизионные станции. Разрешение видео составляет 1280×720 пикселей, что является стандартом для трансляции спортивных событий. Эта гистограмма показана на рисунок 8. На рисунке видно, что размер головы варьируется от 15 до 35 пикселей. Причиной этого изменения является камера, которая в некоторые моменты увеличивает текущую активность на площадке, чтобы более четко показывать зрителям (например, прыжок за мячом после пропущенного удара, когда большое количество игроков находится ниже корзины).

В работах [22, 28] создание пирамиды изображений и карты объектов осуществляется через равные дискретные интервалы, поскольку процент изображения, захватываемого требуемым объектом, неизвестен. Поскольку целью данного исследования является обнаружение игрока на изображениях, транслируемых телевизионными станциями, и поскольку вещание в основном осуществляется в формате HD, можно применять различные выборки для достижения лучшего времени выполнения. Если мы применим реализацию пирамиды объектов из [22, 28] к изображению с разрешением 1280×720 пикселей, мы получим 52-уровневую пирамиду. Следовательно, размер головы – это размер области, которая будет использоваться при обучении (рисунок 6в). Затем, основываясь на гистограмме на рисунке 9, мы можем сделать вывод, что нам не нужно столько уровней в пирамиде, как в [22, 28], а только подмножество. Другими словами, вместо 52 уровней мы использовали только 12 уровней. Такая выборка не повлияла на результат обнаружения (количество обнаруженных объектов осталось прежним), но она повлияла на время выполнения. Среднее время выполнения обнаружения для изображения, полученного в баскетбольном матче, было уменьшено с 267,5 секунд до 185,8 секунд, что означает экономию времени почти на 30%. Измерение проводилось на ПК с процессором i7 и 8 ГБ оперативной памяти. Сокращение времени не больше, потому что мы использовали первые 12 карт признаков, которые являются самыми большими и, естественно, требуют больше времени для обработки.

|  |
| --- |
| Описание: C:\Users\Wicher\Pictures\piramid.png |
| Рисунок 8 – Пирамида изображения, созданная сглаживанием и подвыборкой, а также создание пирамиды объектов путем расчета карты объектов на каждом уровне пирамиды изображений |

|  |
| --- |
|  |
| Рисунок 9 – Гистограмма размеров головы игроков в пикселях |

После создания пирамиды изображений и карты объектов мы можем наложить на них модель. Модель, которая была обучена для каждой части тела (создание модели обсуждалось в предыдущем разделе и выражается выражением (1)). Используя ее, мы получаем оценку, которая говорит нам, насколько вероятно, что на наблюдаемом месте и масштабе в пределах изображения у нас есть необходимая часть. Чтобы получить оценку для конкретной конфигурации деталей, необходимо определить функцию совместимости для типов деталей:

|  |  |
| --- | --- |
|  | (2) |

Параметр поддерживает определенные назначения типов для части , в то время как парный параметр поддерживает конкретные совместные вхождения типов частей. С мы можем пометить реляционный граф, ребра которого определяют, какие пары частей должны иметь согласованные отношения. Этот реляционный граф показан на рисунке 6г. Теперь мы можем написать выражение для полной оценки, связанной с конфигурацией типов деталей и позиций:

|  |  |
| --- | --- |
|  | (3) |

Здесь – вектор признаков (HOG), полученный из местоположения пикселя на изображении .

Выражение представляет относительное местоположение части относительно части . Первая сумма в (3) – это модель внешнего вида, которая вычисляет локальную оценку для размещения шаблона для части , настроенного для типа , в местоположении . Вторая сумма в (3) управляет относительным размещением частей и , переключаясь между коллекциями пружин. Каждая пружина создается для определенной пары типов .

Таким образом, это максимизация из выражения (3) по параметрам и. Когда реляционный граф представляет дерево, максимизация может быть эффективно выполнена с помощью динамического программирования [22]. Сообщение, которое часть передает своей родительской части , может быть вычислено следующими выражениями:

|  |  |
| --- | --- |
|  | (4) |
|  | (5) |

В (4) представляет набор потомков части в . Выражение (4) вычисляет локальную оценку части для всех положений пикселя и для всех типов . Это делается путем сбора сообщений, отправленных потомками наблюдаемой части . Выражение (5) для каждого возможного местоположения и типа части вычисляет наилучшее местоположение оценки и тип ее дочерней части . Когда все сообщения передаются в корневую часть , вычисленная оценка представляет наилучшую возможную конфигурацию частей.

**3.3.4 Удаление ложноположительных обнаружений**

Ложноположительные обнаружения – это те объекты, которые ошибочно классифицируются как обязательные объекты. В нашем исследовании в качестве ложных срабатываний мы обычно регистрируем людей в зале и судей на площадке. Основной причиной их обнаружения является их сходство с требуемыми объектами. Целью данного исследования является определение позиции игрока на площадке. Позиции людей в зале и судей не имеют значения, поэтому мы можем исключить их из дальнейшей обработки.

**3.3.5 Удаление зала**

Лица в зале могут быть исключены из дальнейшей обработки на основании того, что они не находятся на площадке. Вот почему нам нужно определить границы площадки. Чтобы определить площадку, мы сначала применяем алгоритм обнаружения краев Canny [10]. Результат применения этого алгоритма показан на рисунке 10. После этого мы определяем горизонтальную линию с наибольшим количеством белых пикселей. Эта линия представляет горизонтальную границу площадки. Следующим шагом является определение максимальной вертикальной линии, идущей от нижнего края изображения к горизонтальной границе. Наклон этой линии можно определить по наклону горизонтальной линии. Это значительно ускоряет время обнаружения вертикальной границы. Площадка определяется этими двумя линиями и их точкой пересечения (рисунок 10в).

Когда мы обнаруживаем область, которая представляет площадку, мы можем ограничить объекты, чьи лодыжки находятся за пределами найденной области. Это может быть сделано, потому что мы использовали детектирование по частям. После этого шага мы удалили все объекты из зала, оставив только объекты на площадке.

|  |
| --- |
| Описание: C:\Users\Wicher\Pictures\vlcsnap-2019-06-27-00h41m44s558.png |
| (а) Один кадр из баскетбольного матча |
| Описание: C:\Users\Wicher\Pictures\canny.png |
| (б) Применение алгоритма обнаружения краев Canny |
| Описание: C:\Users\Wicher\Pictures\canny_line.png |
| (в) Обнаружение границ площадки |
| Рисунок 10 – Обнаружение границ баскетбольной площадки с использованием алгоритма обнаружения краев Канни |

**3.3.6 Удаление судей**

Чтобы удалить судей, мы вычислили цветовую гистограмму H-компонента из цветовой системы HSV, как в [29]. Однако в [29] рассчитываются цветные гистограммы баскетболистов, и с их помощью определяется, представляет ли наблюдаемый объект игрок или что-то еще (например, рефери). Используя эти гистограммы, мы также удаляем объекты из зала. Этот подход имеет недостаток. Он должен быть переобучен для каждой новой игры, потому что у каждой команды свой цвет формы. В нашем исследовании мы учли тот факт, что все судьи во всей лиге имеют одинаковый цвет футболки. Именно поэтому однажды освоенный алгоритм не нуждается в обучении снова. В нашей работе мы использовали меру сходства Quadratic-Chi между гистограммами, которая показывает лучшие результаты с точки зрения изменений света и деформаций формы. Поскольку в [30] использовался алгоритм обнаружения, который автоматически определяет ключевые точки в тренировочном процессе, невозможно сказать, что является рукой, ногой, головой и т.д. Именно поэтому цветовая гистограмма рассчитывается по всему объекту. В нашем подходе каждая часть определяется отдельно. Цветовая гистограмма рассчитана по шести частям, которые представляют ноги обнаруженного объекта (детали 12, 13, 14, 24, 25 и 26 на рисунке 6г). Таким образом, мы достигли большей точности благодаря значительно меньшей базе данных, предназначенной для обработки.

По сходству гистограммы мы определяем объекты, которые представляют судей. Эти объекты затем исключаются из дальнейшей обработки.

**3.3.7 Разделение игроков на команды**

Когда процесс обнаружения и устранения ложных срабатываний завершен, все обнаруженные объекты представляют баскетболистов. Среди них есть игроки обеих команд. Наш следующий шаг – разделение игроков на две команды в соответствии с цветом их майки. Этот процесс показан на рисунке 11. Первым шагом является выбор областей для кластеризации (рисунок 11а). На самом деле это выбор из тех частей, которые покрыты одеждой. В случае баскетболистов этими частями являются 3, 8, 9, 10, 11, 12, 15, 20, 21, 22, 23 и 24, которые показаны на рисунке 6г. На основе серии мы создали одну область, которая будет использоваться для дальнейшей обработки. Затем эта область преобразуется в цветовое пространство HSV по двум причинам. Первым является тот факт, что компонент H (Hue) этой цветовой системы представляет цвет. Определяя диапазон, который представляет кожу игрока и площадку, эти пиксели могут быть легко удалены из дальнейшей обработки (рисунок 11б). Пиксели кожи удалены, потому что мы хотим разделить игроков по цвету майки, а не по цвету кожи. Другая причина выбора цветового пространства HSV состоит в том, что у каждой команды в лиге есть два вида формы. Один набор «светлый», а другой «темный». Оба набора созданы в соответствии с цветами клуба, но с предыдущим ограничением. В любой конкретной игре команды находятся в разных видах формы (одна команда находится в более светлом наборе, а другая в более темном). Если мы наблюдаем компонент насыщения S моделей HSV, светлый набор имеет низкий уровень насыщенности, в то время как темный набор имеет высокий уровень насыщения. Это именно та характеристика, которую мы использовали при расчете гистограмм и групповой кластеризации. Гистограмма рассчитывается с использованием значения насыщенности для этих пикселей из области вокруг формы игрока, за исключением цвета кожи или площадки. Мы рассчитываем 100-битную гистограмму, из которой выбираем пять пиков с наибольшим значением (рис. 11в). Аналогичный принцип, но для определения цвета площадки был использован в [27], с той разницей, что там использовали компонент H и один или два пика. Причина извлечения пиков в нашей работе заключается в том, что изображения не имеют одинаковый размер, потому что неизвестно, сколько пикселей будет исключено из дальнейшей обработки, поскольку они представляют либо площадку, либо кожу. Пять пиков от гистограммы каждого игрока входят в процесс кластеризации по принципу k-средних. Объекты, принадлежащие первому кластеру, являются игроками первой команды, а объекты, принадлежащие другому кластеру, представляют игроков второй команды. При использовании этого подхода нет необходимости переучивать алгоритм для новых команд и цветов формы. Чтобы определить, какой кластер представляет какую команду, нам просто нужно знать, какая команда в наблюдаемой игре носит более светлый набор, и наоборот. Разделение игроков показано на рисунке 11г.

|  |
| --- |
| Описание: C:\Users\Wicher\Pictures\comands.png |
| а) Выбор области для кластеризации |
| Описание: C:\Users\Wicher\Pictures\comands2.png |
| (б) Удаление пикселей игровой площадки и кожи |
|  |
| (в) Расчет гистограммы и выбор доминирующих выборок |
| Описание: C:\Users\Wicher\Pictures\labeled_comands.png |
| (г) Результат процесса кластеризации |
| Рисунок 11 – Разделение игроков по составу команды |

**3.3.8 Определение позиции игрока**

Когда все игроки обнаружены и разделены на команды, мы хотим определить их текущую позицию на площадке. Влияние точности удара с разных позиций на командную игру дано в исследовании [25]. Разделение площадки может быть сделано разными способами. Если мы изучим специализированную литературу, связанную с баскетболом, то заметим, что общего решения проблемы не существует. Причина может быть найдена в том, что производительность игрока, как правило, наблюдается в отношении его позиции. Таким образом, мы могли бы создать подразделение площадки для центрального игрока, разыгрывающего и т.д. В нашем исследовании мы использовали разделение, которое применяется в большинстве баскетбольных лиг: площадка делится на 11 областей. Эти области показаны на рисунке 8г. На рисунке обозначены шесть областей, но области 1, 2, 3, 4 и 6 содержат подрайон, из которого успешные выстрелы считаются двумя точками, и подрайон, из которых успешные выстрелы считаются тремя точками. Поэтому эти районы учитываются дважды. В некоторых лигах область 5 разделена на четыре подрайона. Причина в том, что под корзиной у нас большая часть активности. Дополнительное деление представляет собой попытку записать все эти действия.

Для того чтобы определять позиции игроков, мы расширили определение площадки на основе алгоритма обнаружения ребер Canny. Помимо линий, которые определяют площадку, мы обнаружили линии, которые определяют краску (трехсекундную зону). Первым шагом является определение двух более длинных границ краски, которые основаны на том факте, что они почти параллельны горизонтальной границе площадки. Затем мы нашли вертикальную границу краски (линия штрафного броска), которая почти параллельна вертикальной границе площадки. Обнаружение корта и краски показано на рисунке 12б. После этого шага у нас есть пять точек с точным расположением на изображении. Это точки в четырех вершинах краски, и точка, которая лежит на пересечении горизонтальной и вертикальной границ площадки. Взяв фактический размер площадки в лиге НБА вместе с этими пятью точками, мы можем затем определить функцию, которая представляет пространственное преобразование. Это пространственное преобразование отображает точку от изображения, скоординированного до точки в реальной системе координат площадки. Пара точек имеет отношение , где представляет матрицу преобразования 3×3. Если точки реальной системы находятся в одной плоскости, как в нашем случае, называется матрицей гомографии. При определении этого пространственного преобразования мы фактически задали функцию, которая позволит всем точкам на изображении определить свое точное положение на площадке. Эта функция поможет нам определить позиции игрока на площадке.

Прежде чем указывать положение игрока на площадке, необходимо определить его положение на изображении. В предыдущей части этого текста мы объяснили, как выполнить обнаружение игрока. Однако теперь нам нужна одна точка на игрока, которая представляет точную позицию на площадке. Наблюдая некоторые совпадения, мы пришли к выводу, что эта точка должна лежать между ногами игрока. Мы также определили его координату этой точки прямоугольниками, которые отмечают целых игроков (рисунок 11г). Наша координата расположена между двумя координатами , которые представляют левый и правый края прямоугольника. Координата рассчитывается с использованием прямоугольников, которые представляют лодыжки игрока (области 14 и 26 на рисунке 12г). Требуемая координата находится между нижними координатами этих прямоугольников. Пример обнаруженных игроков и их позиций на площадке показан на рисунке 12.

|  |  |
| --- | --- |
| Описание: C:\Users\Wicher\Pictures\vlcsnap-2019-06-27-00h41m44s558.png | Описание: C:\Users\Wicher\Pictures\line.png |
| (а) Кадр из баскетбольного матча | (б) Обнаружение границ площадки и краски |
| Описание: C:\Users\Wicher\Pictures\labeled_comands.png |  |
| (в) Обнаружение и разделение игроков на команды | (г) Отображение позиции игрока |
| Рисунок 12 – Определение положения игрока и его отображение на площадке | |

**3.4 Разработка алгоритмов трекинга объектов на видеоданных с диагонального ракурса камеры**

В отслеживании объектов модель наблюдения и стратегия отслеживания являются двумя ключевыми компонентами. Для отслеживания нескольких игроков в спортивном видео требуется еще больше усилий для построения модели наблюдения и разработки стратегии отслеживания.

**3.4.1 Модель наблюдения**

В связи с быстрым развитием исследований по отслеживанию объектов в настоящий момент используются различные виды наблюдений для отслеживания объектов, такие как характерные точки, шаблоны и движущиеся области. В последнее время быстрое развитие методов обнаружения объектов привело к появлению многих перпективных методов обнаружения объектов определенных классов, например, лиц и пешеходов. Эти детекторы объектов предоставляют хорошие модели наблюдения для алгоритмов отслеживания на основе обнаружения. За счет обеспечения конфиденциальности вывода детекторов для мягкого решения, их можно использовать при инициализации, росте и завершении траектории объекта. Однако детекторы далеки от совершенства. Пропуски цели, ложные тревоги и неточные реакции обнаружения – это три основные проблемы для многих алгоритмов отслеживания объектов на основе обнаружения. Чтобы улучшить модель наблюдения, многие методы объединяют несколько сигналов одновременно, таких как внешний вид объекта, форма и движение. Оценка эффективности различных сигналов и способ их организации, однако, все еще требуют дальнейшего анализа.

**3.4.2 Стратегия отслеживания**

После того как установлено наблюдения за объектами, появляется проблема, заключающаяся в том, как использовать их в системе отслеживания. Как правило, есть две разные стратегии. Первая – использовать их с точки зрения вероятностного вывода. Фильтр Калмана является ранним представительным методом такой стратегии. В последнее время, основанные на последовательной выборке Монте-Карло, частичные фильтры привлекли большое внимание из-за их простоты, универсальности и масштабируемости в широком спектре сложных приложений. Одно успешное применение частичного фильтра – отслеживание одиночного объекта (SOT). Тем не менее, использование единой модели наблюдения для всех объектов все еще не может их хорошо различить, особенно когда представлены окклюзии. Для дальнейшего решения проблемы окклюзии нескольких объектов методы отслеживания на основе глобальных ассоциаций предоставляют еще одну новую стратегию с точки зрения детерминированной оптимизации. С учетом откликов на обнаружение эти методы пытаются связать их глобально для решения проблемы взаимодействия нескольких объектов путем адаптации определенного вида алгоритма оптимизации, такого как алгоритм Витерби [17], линейное программирование [18], поток минимальных затрат [19] и алгоритм Hungary [20]. Поскольку информация обо всех кадрах доступна до отслеживания, эти методы теоретически могут достичь глобального оптимума. Однако основным недостатком являются большое количество вычислений и временные задержки, которые ограничивают их использование для приложений реального времени.

**3.4.3 Отслеживание объекта как последовательная байесовская оценка**

При обозначении последовательности состояния объекта как , что обычно считается марковским процессом, и последовательности наблюдения как , отслеживание объекта часто формализуется в виде задачи последовательной байесовской оценки с помощью двухэтапной рекурсии Prediction and Update следующим образом:

|  |  |
| --- | --- |
|  | (6) |
|  | (7) |

где – динамическая модель системы, которая отражает стратегию отслеживания, –модель наблюдения, которая дает вероятность наблюдения в пространстве состояний. Здесь мы можем далее понять важность модели наблюдения и стратегии отслеживания для системы отслеживания. Если мы хотим улучшить производительность системы отслеживания, оба эти компонента должны быть тщательно разработаны.

На практике распределение фильтрации может быть аппроксимировано набором взвешенных частиц в последовательной структуре Монте-Карло (также известной как фильтр частиц):

|  |  |
| --- | --- |
|  | (8) |

где – общее количество частиц, обозначает дельта-дираковскую функцию в точке .

Эта формулировка стала популярна для SOT и показала большие успехи в отслеживании таких объектов, как рука, лицо и голова. Однако при применении его к MOT возникают как минимум две проблемы. Одна из них заключается в том, как назначить соответствующие наблюдения каждому объекту, когда несколько объектов взаимодействуют друг с другом. Мы называем это проблемой назначения наблюдения, которая обычно вызывает переключение идентификатора при отслеживании двух соседних объектов. Другая проблема состоит в том, чтобы продолжить отслеживание, когда наблюдения не могут быть собраны для одного объекта из-за окклюзии или резких движений. Мы называем это проблемой отсутствия наблюдений, которая обычно вызывает фрагментацию на одной траектории объекта. Эти две проблемы намного более распространены в отслеживании нескольких игроков из-за их очень динамичных и высоко интерактивных движений.

**3.4.4 Отслеживание объектов как двойной режим двухстороннего байесовского вывода**

Чтобы расширить вышеуказанную модель до МОТ, обычно используются две формулировки. Первая заключается в инициализации новой дорожки для каждого объекта и отслеживании их независимо. Другая заключается в том, чтобы рассматривать все объекты в целом и отслеживать их глобально. С одной стороны, в то время как первая формулировка проста для понимания и реализации, все еще очень трудно решить проблему назначения наблюдения. С другой стороны, последняя формулировка может дать более общее представление, но она часто делает проблему неразрешимой или требует сложных вычислений.

В этой статье мы следуем стратегии «разделяй и властвуй», чтобы сформулировать проблему MOT как объединение отслеживания одиночного изолированного объекта и отслеживания множественных скрытых объектов. Обозначая как набор состояний всех объектов за время , мы разделяем набор объектов на два подмножества в соответствии с их пространственными отношениями: набор изолированных объектов и набор скрытых объектов . Для каждого объекта мы отслеживаем его независимо с помощью общей модели наблюдения, которая предназначена для того, чтобы отличать их от фона. Что касается объектов , мы рассматриваем каждый кластер окклюзии в целом и отслеживаем их одновременно с помощью специальной модели наблюдения, которая предназначена для различения разных игроков. Предположим, что в один закрытый кластер в момент входит закрытых объектов, – состояния объектов, – подтверждающие наблюдения. Соответствующее распределение фильтров может быть аппроксимировано как:

|  |  |
| --- | --- |
|  | (9) |

Теперь проблема назначения наблюдений может быть естественным образом решена путем сбора модели-кандидата, поддерживающей наблюдения для каждого объекта. Далее мы рассмотрим проблему отсутствия наблюдений, которая решается путем введения концепции сглаживания в задачу отслеживания.

Вышеупомянутая формулировка задачи отслеживания также известна как прямая фильтрация в задаче оценки. Это соответствует оценке распределения текущего состояния системы на основе наблюдений, полученных до текущего времени. Эта формулировка делает систему отслеживания быстро адаптирующейся к изменениям и легко внедряемой. Основная проблема, однако, заключается в том, что отсутствует чувствительность к временному наблюдению. Если один объект не может получать наблюдения в нескольких кадрах, процесс прямой фильтрации потеряет его, даже если этот объект можно будет позже наблюдать. В этой работе мы преодолеваем данную проблему, вводя процедуру обратного сглаживания в задачу отслеживания. Основная идея процедуры обратного сглаживания проста: если система не может получить достаточное количество наблюдений для одного объекта в текущем кадре, она будет ожидать и повторно отслеживать объект, когда позднее будет получено достаточное количество наблюдений. Сглаживание в обратном направлении является еще одной важной концепцией в задаче оценки, которая соответствует оценке распределения состояния в конкретное время с учетом всех наблюдений до более позднего времени . Для одного объекта это равносильно оценке распределения . Теоретически, это предельное распределение может быть получено непосредственно путем интегрирования переменных из совместного распределения , что может быть представлено как:

|  |  |
| --- | --- |
|  | (10) |

Для оценки этого распределения можно использовать несколько стратегий [29], если существует только один объект. В нашем случае, однако, распределения нескольких ненаблюдаемых объектов должны оцениваться одновременно. При обозначении буферизованного ненаблюдаемого набора объектов в процедуре прямой фильтрации как , процесс обратного сглаживания можно сформулировать как:

|  |  |
| --- | --- |
|  | (11) |

Вместо того, чтобы моделировать это распределение с помощью частичной дискретизации, мы напрямую находим его максимальную апостериорную вероятность посредством процесса сопоставления между распределением и . На основе результата сопоставления траектории буферизованных объектов были успешно отслежены, связаны и сглажены с использованием соответствующего наблюдения.

К настоящему времени мы разработали нашу формулировку задачи. Поскольку она дает единую байесовскую формулировку проблем SOT и MOT посредством прямой фильтрации и обратного сглаживания, мы назвали такую схему - режим двухстороннего байесовского вывода (рисунок 14).

|  |
| --- |
|  |
| Рисунок 14 – Блок-схема предлагаемой системы отслеживания игроков на площадке |

Предложенная схема будет использована для проведения теоретических и экспериментальных работ на следующих этапах НИОКР.

**4. ЗАКЛЮЧЕНИЕ**

Основные результаты НИОКР на данном этапе можно сформулировать следующим образом:

* Для разработки системы спортивной видеоаналитики на первом этапе работ была приобретена антивандальная купольная камера GV-EVD3100 - разработанная в соответствии со стандартами IK10+ и IP67, что позволяет применять данную купольную камеру для подобных систем, без риска выведения из строя системы при попадании спортивными снарядами, такими как баскетбольный мяч. Камера оснащена автоматическим ИК-фильтром и ИК-светодиодами, а также светочувствительной 1/2.8” CMOS-матрицей с прогрессивной разверткой, способной обеспечить цветное изображение в почти полной темноте. Регуляция положения объектива по 3 осям позволяет размещать данную камеры как на потолке спортивного зала, так и на стене, и получать ракурсы съемки, необходимые для работы систем спортивной видеоаналитики.
* Для обучения алгоритмов искусственного интеллекта была проведена съемка реальной игры баскетбольных команд в спортивном зале ЯрГУ им. П.Г. Демидова из которой была подготовлена база изображений, состоящая из 1500 изображений баскетболистов, с разрешением от 50x100 пикселей до 200x400 пикселей.
* Для тестирования алгоритмов спортивной видеоаналитики была подготовлена тестовая база изображений баскетболистов, состоящая из 500 изображений, с разрешением от 50x100 пикселей до 200x400 пикселей.
* Был разработан алгоритм детектирования объектов на видеоданных с диагонального ракурса камеры, содержащий следующие основные этапы: обнаружение объектов, создание модели игрока, модель обучения, удаление ложноположительных обнаружений, удаление зала, удаление судей, разделение игроков на команды, определение позиции игрока.
* Разработан алгоритм трекинга (отслеживания) объектов на видеоданных с диагонального ракурса камеры. Предложена блок-схема набора алгоритмов для отслеживания игроков. Она дает единую байесовскую формулировку проблем SOT и MOT посредством прямой фильтрации и обратного сглаживания (режим двухстороннего байесовского вывода).

Областями применения разработанных алгоритмов являются комплексы спортивной видеоаналитики, которые будут использованы спортивными клубами, многофункциональными спортивными комплексами.

Работы по данному этапу НИОКР были выполнены в полном объеме в соответствии с техническим заданием и календарным планом. Полученные результаты будут использованы на следующих этапах НИОКР и в коммерческой деятельности ООО «Цифровые решения».

# 5. СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных, - М.: ДМК Пресс, 2015.
2. Бринк Х., Ричардс Д., Феверолф М. Машинное обучение. – Спб.: Питер, 2017.
3. Концепция построения и развития аппаратно-программного комплекса «Безопасный город» // Распоряжение Правительства Российской Федерации от 3.12.2014 № 2446-р.
4. Ko T. A Survey on behavior analysis in video surveillance for homeland security application // 37th IEEE Applied Imagery Pattern Recognition Workshop. AIPR, 2008. P. 1–8.
5. Zhao W., Chellappa R., Phillips P., Rosenfeld A. Face recognition: A literature survey // ACM Computing Surveys (CSUR). 2003. V. 35, № 4. P. 399–458.
6. Форсайт Д.А., Понс Д. Компьютерное зрение. Современный подход. – М.: Вильямс, 2004. – 928 с.
7. Woodward J.D., Orlans N.M., Higgins P.T. Biometrics: Identity Assurance in the Information Age. – McGraw-Hill Osborne Media, 2002. – 416 p.
8. Brey P. Ethical Aspects of Face Recognition Systems in Public Places // Journal of Information, Communication & Ethics in Society. 2004.  
   P. 97–109.
9. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2005. – 621 с.
10. Методы компьютерной обработки изображений / Под ред. В.А. Сойфера. – М.: Физматлит, 2001. – 784 с.
11. Журавлев Ю.И., Рязанов В.В., Сенько О.В. Распознавание. Математические методы. Программная система. Практические применения. – М.: Фазис, 2005. – 159 с.
12. Sagonas C., Tzimiropoulos G., Zafeiriou S., Pantic M. 300 Faces in-the-Wild Challenge: The first facial landmark localization Challenge // Proc. International Conference on Computer Vision workshop on 300 Faces in-the-Wild Challenge, 2013. P. 397–403.
13. Хайкин С. Нейронные сети: Полный курс. Пер. с англ. Н.Н. Куссуль, А.Ю. Шелестова. 2-е изд. – М.: Издательский дом Вильямс, 2008.
14. Bishop C.M. Pattern Recognition and Machine Learning, Springer, 2006.
15. Rokach, L. Ensemble-based classifiers // Artif Intell Rev (2010) 33: 1.
16. Alahi A, Boursier Y, Jacques L, Vandergheynst P (2009) Sport players detection and tracking with a mixed network of planar and omnidirectional cameras. In: ACM/IEEE International conference on distributed smart cameras, pp 1–8 Multimed Tools Appl (2014) 72:2741–2767 2765.
17. Amit Y, Trouve A (2007) POP: patchwork of parts models for object recognition. Int J Comput Vis 75(2):267–282.
18. Bar-Hillel A, Weinshall D (2008) Efficient learning of relational object class models. Int J Comput Vis 77(1):175–198.
19. Bourdev L, Malik J (2009) Poselets: Body part detectors trained using 3d human pose annotations. In: International conference on computer vision (ICCV), pp 1365–1372.
20. Chang MH, Tien MC, Wu JL (2009) WOW: wild open warning for broadcast basketball video based on player trajectory. In: ACM International conference on multimedia, pp 821–824.
21. Dalal N, Triggs B (2005) Histogram of oriented gradients for human detection. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 886–893.
22. Felzenszwalb P, Girshick R, McAllester D, Ramanan D (2010) Object detection with discriminatively trained part based models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645.
23. Felzenszwalb P, Huttenlocher D (2005) Pictorial structures for object recognition. Int J Comput Vis 61(1):55–79.
24. Fergus V, Perona P, Ziserman A (2003) Object class recognition by unsupervised scale-invariant learning. In: IEEE computer society conference on computer vision and pattern recognition.
25. Ivankovic Z, Rackovic M, Markoski B, Radosav D, Ivkovic M (2010) Appliance of neural networks in basketball scouting. Acta Polytech Hungarica 7(4):167–180.
26. Szeliski R (2010) Computer vision: algorithms and applications. Springer, New York.
27. Wu L, Gong Y, Yuan X, Zhang X, Cao L (2012) Semantic aware sport image resizing jointly using seam carving and warping. Multimed Tools Appl. doi:10.1007/s11042-012-1002-7 .
28. Yang Y, Ramanan D (2011) Articulated pose estimation using flexible mixture of parts. Comput Vis Pattern Recogn (CVPR) 1385–1392.
29. Lifang W, XiuliH, Hao C, Wei S (2007) Shot segmentation and classification in basketball videos. In: Intelligent information hiding and multimedia signal processing, pp 539–542.
30. Lu WL, Ting JA,Murphy KP, Little JJ (2011) Identifying players in broadcast sports videos using conditional random fields. In: IEEE Conference on computer vision and pattern recognition (CVPR), pp 3249–3256.