На правах рукописи

Гиля-Зетинов Александр Александрович

РАЗРАБОТКА МЕТОДОВ И АЛГОРИТМОВ СОПРОВОЖДЕНИЯ ОБЪЕКТОВ В СИСТЕМАХ МАШИННОГО ЗРЕНИЯ

Специальность 1.2.2 —
«Математическое моделирование, численные методы и комплексы программ»

Автореферат

диссертации на соискание учёной степени
кандидата физико-математических наук

Долгопрудный — 2022

Работа выполнена в федеральном государственном автономном образовательном учреждении высшего образования «Московский физико-технический институт (национальный исследовательский университет)».

Научный руководитель: **Северов Дмитрий Станиславович** - кандидат физико-математических наук

Ведущая организация: Федеральное государственное бюджетное учреждение высшего профессионального образования «Московский автомобильно-дорожный государственный технический университет (МАДИ)»

Защита состоится XX ноября 2022 г. в 11 часов на заседании диссертационного совета ФЭФМ.05.13.18.001 на базе ФГАОУ ВО Московский физико-технический институт (национального исследовательского университета) по адресу: 141701, Московская область, г. Долгопрудный, Институтский переулок, д.9.

С диссертацией можно ознакомиться в библиотеке Федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)»

Работа представлена «___ » _____ 2021 г. в Аттестационную комиссию Федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)» для рассмотрения советом по защите диссертаций на соискание ученой степени кандидата наук в соответствии с п. 3.1 ст. 4 Федерального закона «О науке и государственной научно-технической политике».

Общая характеристика работы

Актуальность темы. В данной работе описывается ряд новых подходов к задаче отслеживания фигур людей , возникающей при разработке и применении систем машинного зрения для анализа тактических сцен с большим количеством участников. Такие системы в последние годы активно применяются в различных областях. Примерами могут служить системы общественной и промышленной безопасности, решения для автоматического анализа спортивных трансляций, различные решения в составе систем "Умный город".

Актуальность темы исследования.

Одним из важных этапов анализа видео сцены с большим количеством участников является переход от набора обнаруженных в каждом кадре участников к построению траекторий движения участников в пространстве изображения и пространстве сцены. Различные вариации задачи сопровождения, также именуемой задачей множественного отслеживания объектов (MOT — multiple object tracking), известны давно. Но в связи с распространением нейронных сетей как основного алгоритма в машинном зрении, данная задача вновь стала вызывать повышенный интерес у исследователей. Это связано с несколькими факторами.

Во-первых, существует множество дополнительных данных, которые может предоставить нейросетевой алгоритм помимо предполагаемых координат объекта и которые можно использовать для увеличения качества отслеживания.

Примером таких данных, формируемых нейросетью, может послужить вектор, описывающий позу человека, визуальные характеристики или предполагаемое направление движения.

Во-вторых, вывод нейронной сети имеет свою специфику ошибок, отличающихся от классических моделей — а используемая модель ошибок играет важную роль в разработке алгоритма.

В-третьих, в последнее время распространены исследования о применимости нейронных сетей в роли алгоритмов отслеживания. К этому привело появление новых архитектур сетей — графовых нейронных сетей, а также сетей, основанных на операции внимания.

Таким образом, задача разработки новых подходов к решению проблемы MOT, предназначенных для совместного использования с нейросетевыми алгоритмами обнаружения отметок, является актуальной.

Целью данной работы является разработка и реализация методов решения ряда промежуточных задач в системах машинного зрения, связанных с построением траекторий объектов – людей в высоконагруженных сценах. Для ее достижения решаются следующие **задачи**:

1. Выбор и реализация алгоритма анализа одиночного изображения для поиска людей и координат их скелетной модели, для получения исходных данных для дальнейшей обработки.
2. Разработка модели для решения задачи множественного отслеживания объектов - фигур людей в сценах с большим количеством объектов и нейросетевыми технологиями формирования отметок (обнаружений) в отдельных кадрах.
3. Разработка алгоритмов отслеживания для объектов в пространстве координат кадра, представленных скелетной моделью.
4. Разработка алгоритма перевода координат наблюдаемых объектов из пространства кадра в пространство сцены, с учетом оптических искажений камер.
5. Разработка стохастического алгоритма отслеживания объектов с сопутствующей фильтрацией координат и построением траектории в высоконагруженных сценах.
6. Оценка качества и скорости работы реализованных алгоритмов отслеживания в зависимости от значений их параметров с использованием размеченных наборов данных.
7. Практическое применение разработанных алгоритмов и их программных реализаций для систем общественной и промышленной безопасности, решений для автоматического анализа спортивных трансляций, различных решений в составе систем "Умный город".

Основные положения, выносимые на защиту:

1. Модель для решения задачи отслеживания объектов - фигур людей в сценах с большим количеством объектов и нейросетевыми технологиями формирования отметок (обнаружений) в отдельных кадрах [1; 2];

2. Новый метод отслеживания объектов, представленных скелетной моделью, на основе поиска максимального парасочетания в двудольном графе (PBBM – Pose-Based Bipartite Matching) [3].
3. Новый метод отслеживания и фильтрации траекторий объектов в пространстве координат сцены с использованием метода Монте-Карло (MCTO – Monte-Carlo Trajectory Optimization) [4].
4. Программная компонента для выделения и отслеживания людей в пространствах координат и сцены по видеоряду на основе вышеперечисленных методов [5; 6].

Научная новизна:

1. Впервые предложен метод отслеживания представленных скелетной моделью объектов, на основе метода двудольного сопоставления с одновременным учетом линейной модели движения, координат суставов и внешних признаков (PBBM – Pose-Based Bipartite Matching).
2. Впервые предложен метод отслеживания и фильтрации траектории в задаче множественного отслеживания объектов на основе поиска набора наиболее правдоподобных управляющих векторов путем стохастической оптимизации в скользящем окне (MCTO – Monte-Carlo Trajectory Optimization).

Практическая значимость . В работе предлагается два новых метода отслеживания с разными областями применимости. Метод отслеживания по скелетной модели PBBM относится к широко распространенной и используемой на практике группе методов, основанных на двудольном сопоставлении. Отличием от других методов в этой группе является вид оптимизируемой функции для весов ребер, использующей одновременно предсказание будущего местоположения, информацию о скелете и цветовую информацию изображения. Интерес данный метод представляет в тех задачах, где уже требуется расчет позы для прикладного применения.

Среди ключевых особенностей, отличающих предложенный стохастический метод MCTO от других методов можно перечислить:

1. промежуточное положение между методами покадровой оптимизации, имеющими нулевую задержку, и глобальными методами, обрабатывающие видео целиком и неприменимые в реальном времени.

Это достигается за счет оконной обработки и наличия фиксированной задержки (как правило, порядка нескольких секунд, но это может быть изменено в зависимости от области применения алгоритма), наличие которой позволяет более точно разрешать окклюзии и восстанавливать продолжительные пропуски.

2. работу в пространстве сцены, а не кадра.
3. оптимизируемая функция может не ограничивается независимыми слагаемыми для пар “трек — обнаружение”, а значит, позволяет более точно учесть специфику обнаружений нейронных сетей.

Некоторые параллели возможно провести с методом объединенного фильтра ассоциации вероятностных данных (MC-JPDAF) — оба этих алгоритма относятся к стохастическим, но имеют разный подход к представлению оптимизируемого пространства.

Научная значимость заключается в разработке данных методов, а также в результатах численных экспериментов на размеченных данных, показывающих качество отслеживания в зависимости от значений параметров.

Практическая значимость подтверждается применимостью комплекса программ, разработанного на основе этих методов на практике. В том числе, практическим применением разработанного комплекса программ при решении задачи мониторинга времени обслуживания авиапассажиров в очередях на территории Московского Аэропорта Шереметьево.

Достоверность полученных результатов обеспечивается расчетом метрик качества на размеченных реальных данных, полученных из открытых источников, а также использованием алгоритмов в прикладных применениях. Результаты находятся в соответствии с результатами, полученными другими авторами.

Апробация работы. Основные результаты работы докладывались на:

1. 4th International Conference on Electrical, Control and Instrumentation engineering (ICECIE), Kuala-Lumpur, Malaysia, 2022
2. Международной конференции Computing Conference 2021, Лондон, Великобритания, 15 июля 2021;
3. XXII Международной конференции «Цифровая обработка сигналов и ее применение DSPA-2020» ИПУ им.Трапезникова";

4. Международной конференция Intelligent Systems Conference (IntelliSys 2020), май 2020, Амстердам, Нидерланды;
5. Международной конференции International Conference on Technology and Entrepreneurship (ICTE), Болонья, Италия, 20-21 апреля 2020

Личный вклад. Разработка алнгоритмов и программная реализация методов отслеживания, выносимых на защиту, а также реализация программного комплекса, были совершены лично автором.

Публикации. Основные результаты по теме диссертации изложены в 5 печатных изданиях, 4 из которых цитируются системой Scopus и рекомендованы ВАК [1—3; 5], 1 — в сборнике докладов рекомендованном ВАК [6]. Еще одна статья принята к публикации в сборнике докладов 4th International Conference on Electrical, Control and Instrumentation engineering (ICECIE), Kuala-Lumpur, Malaysia, 2022 [4].

Содержание работы

Во **введении** обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, приводится обзор научной литературы по изучаемой проблеме, формулируется цель, ставятся задачи работы, сформулированы научная новизна и практическая значимость представляемой работы.

В **первой главе** приведены наиболее общие сведения о существующих подходах к решению задачи отслеживания множества объектов (MOT). В частности описывается место двух новых разработанных автором методов Pose-Based Bipartite Matching (PBBM) и Monte-Carlo Trajectory Optimization (MCTO), предложенных в данной работе на фоне существующих подходов.

Во **второй главе** приводятся сведения о сверточных нейронных сетях, представляющих собой основу для остальной части работы. Первый раздел главы описывает общие приемы анализа изображений с помощью нейронных сетей. Рассматривается три класса задач, представимых через дифференцируемое преобразование изображения - задачи классификации, сегментации и обнаружения объектов. Перечислены особенности применяемых на практике нейронных сетей, основные виды слоев, функций активации. В порядке возникновения приведен обзор основных архитектур нейронных сетей для

обнаружения объектов от первых подходов на основе region proposal, до современных алгоритмов работающих в реальном времени — Faster R-CNN и YOLO.

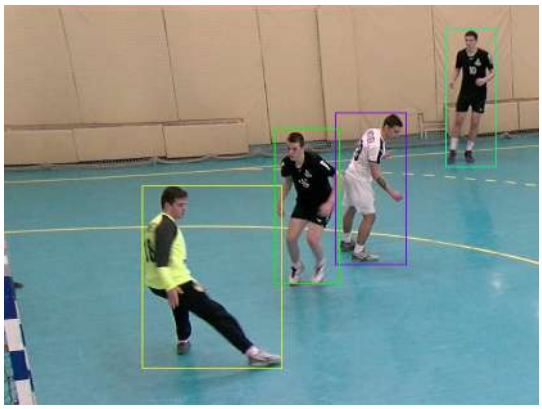

Рис. 1 — Пример работы YOLOv4.

Рис. 2 — Пример работы OpenPose.

Уделено внимание двум архитектурам, использовавшимся для практической реализации программного комплекса. Для выделения объектов используется архитектура обнаружения объектов YOLOv4, обученная на открытом наборе данных MS-COCO и лидирующая по значением метрик на момент разработки (Рис.1).

В задачах, где требуется определение позы используется архитектура OpenPose (Рис. 2).

Третья глава посвящена описанию задачи отслеживания множества объектов (multiple object tracking) и первому предложенному новому методу ее решения на основе двудольного сопоставления с использованием данных о позе — Pose-Based Bipartite Matching (PBBM).

Приводится общая для всех методов модель сцены в задаче. Все измерения и оценки происходят в моменты времени $t = i\Delta t$, $0 \leq i < N_{frame}$, N_{frame} - число кадров. Состояние сцены в данные моменты описывается как совокупность векторов, описывающих активные объекты $\vec{a}_{i,j}$, где $i \in [0, M)$ - номер объекта, j - номер кадра. Содержание вектора $\vec{a}_{i,j}$ - истинного состояния объекта - зависит от используемой модели динамики объектов. Например, в простейшем случае, каждый из векторов $\vec{a}_{i,j}$ может состоять из вектора положения и скорости в неподвижной системе координат сцены. Вектор $\vec{a}_{i,j}$ существует для непрерывного отрезка $[T_i^{(min)}, T_i^{(max)}]$ (времени жизни объекта).

Множество векторов $\vec{a}_{i,j}$ представляет собой скрытое состояние сцены, оцениваемое с помощью алгоритма распознавания объектов. Координаты обнаруженного объекта с индексом j в кадре с номером i обозначаются как $(x_{i,j}^{(det)}, y_{i,j}^{(det)})$, а совокупное число обнаруженных объектов в кадре i как $N_i^{(det)}$. При этом одинаковые индексы j в разные моменты времени могут соответствовать разным истинным объектам сцены. Выходными данными алгоритма отслеживания является множество последовательностей координат объектов. Обозначим их как $(x_{i,j}^{(obj)}, y_{i,j}^{(obj)})$, где i - номер объекта, j - номер кадра.

Помимо модели сцены и ее динамики, необходимо так же описать модель ее наблюдений. Другими словами, необходимо задать вероятностную зависимость:

$$P^{(det)} = P(x_i^{(det)}, y_i^{(det)}, N_i^{(det)} | \vec{a}_{0,i} .. \vec{a}_{M-1,i}) \tag{1}$$

Обычно данная задача решается с использованием предположения о том, что существует некоторое сопоставление между обнаруженными и истинными объектами. Считается, что каждый истинный объект вызывает не более одного обнаруженного, и каждый обнаруженный объект либо соответствует какому-то объекту из истинных, либо является ложным обнаружением не связанным с истинными объектами. В таком случае задача обработки кадра разбивается на два этапа: этап ассоциации между отслеживаемыми объектами и обнаруженными и этап восстановления новых истинных состояний объектов на основе полученных ассоциаций. Это верно для многих методов отслеживания, таких как MHT, MCMCDA, SORT, и пр.

Для реализации метода в данной главе был выбран подход на основе двудольного сопоставления, в котором построение ассоциаций происходит путем поиска паросочетания наибольшего веса в двудольном графе $G(T, D, E)$ (рисунок 3). Возможное оптимальное паросочетание отмечено красными гранями. Активные треки $\mathbf{T}$ хранят в себе всю информацию, необходимую для вычисления весов ребер $\mathbf{E}$.

Вершины в множестве T соответствуют объектам, отслеживаемым на момент кадра i, а вершины в множестве D - обнаруженным объектам в кадре $i + 1$. Взвешенные ребра E соединяют каждую из вершин T со всеми вершинами в D. Вес ребер E отражает степень сходства между ожидаемым состоянием объекта и обнаружением. В случае, если единственными параметрами обнаружения являются координаты, то, как правило, вес ребра находится в

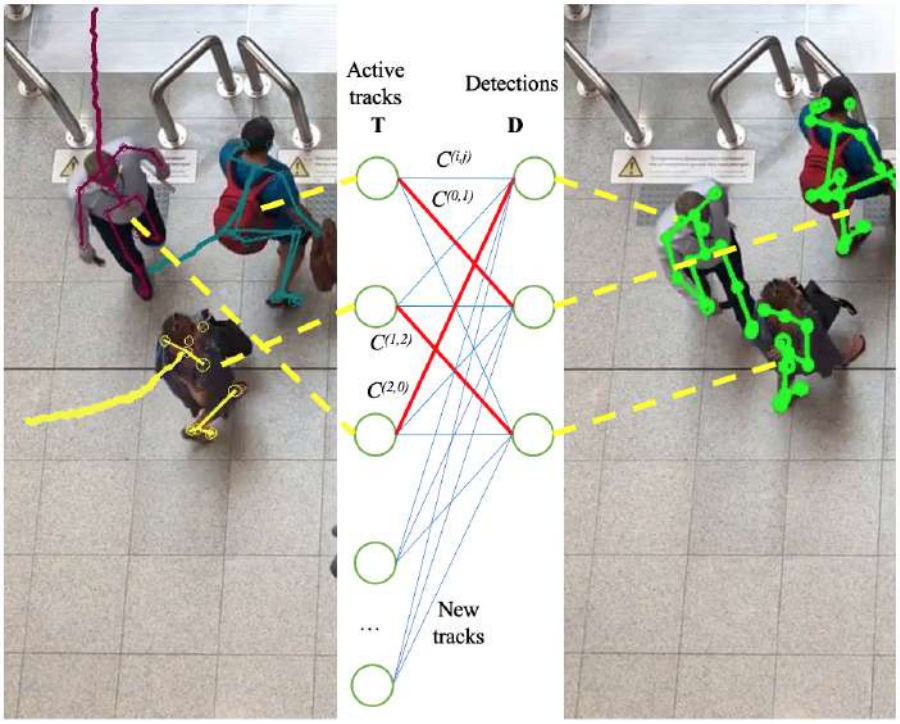

Рис. 3 — Пример паросочетаний между треками и обнаружениями

обратной зависимости от расстояния между ожидаемой позицией объекта и обнаружением, что обусловлено непрерывностью траектории. Также на вес ребра могут влиять и другие оцениваемые параметры, такие как внутренняя структура или цвет объекта.

На основе данного общего подхода предложен новый метод и его реализация для отслеживания скелетных моделей, полученных с помощью OpenPose. Новизна метода заключается в виде используемой функции веса ребра E между i-м треком и j-м обнаруженным объектом:

$$C^{(i,j)} = C^{(i,j)}_{new} + C^{(i,j)}_{center} + C^{(i,j)}_{pose} + C^{(i,j)}_{vis} \tag{2}$$

Тем самым, вес ребра можно разбить на четыре слагаемых:

1. $C^{(i,j)}_{center}$ — слагаемое, пропорциональное расстоянию между предсказанным центром объекта i в следующем кадре и обнаружением j. Оно предназначено для того, чтобы трекер мог сохранять отслеживаемый объект в случае прерывания в обнаружении, предполагая, что скорость движения за этот период существенно не изменилась. Предсказания проводятся на основе линейной регрессии координат объекта в нескольких последних кадрах.

2. $C_{pose}^{(i,j)}$ — слагаемое, пропорциональное степени близости позы между последним ассоциированным скелетом для объекта i и обнаружением j.
3. $C_{vis}^{(i,j)}$ — слагаемое, пропорциональное степени отклонения визуальных признаков обнаружения j от статистики, рассчитанной по истории объекта i, описанное далее.
4. $C_{new}^{(i,j)}$ — штраф за создание нового трека вместо продолжения существующего. Зависимость $C_{new}^{(i,j)}$ от экранных координат для позы j может быть использована для моделирования областей кадра, где наиболее вероятно появление нового объекта, например таких как границы кадра или наблюдаемые выходы.

$C_{vis}^{(i,j)}$ - это слагаемое, отвечающее за сопоставление визуальных признаков между отслеживаемыми людьми и новыми обнаружениями. В связанных работах для этого обычно используется дополнительная сверточную сеть, такую, как VGG либо ResNet. Предлагается вместо этого использовать статистику цвета, полученную из пикселей изображения в окрестностях обнаруженных суставов. Для этого, сначала данные скелетной модели используются для аппроксимации области изображения, содержащей человека, с помощью простых геометрических фигур. Затем для каждой фигуры вычисляются пространственные характеристики распределения цветов внутри этой формы. После чего эти характеристики используются для вычисления уже временной статистики, на основе которой вычисляется весовая функция.

Для валидации алгоритма использовались открытые наборы данных PoseTrack 2017 и 2018 и предназначенных для определения качества решения трех задач: оценка позы в отдельном кадре, оценка позы на видео и отслеживание позы на основе видео.Оценка происходит с помощью метрики средней точности для каждой ключевой точки (mAP), и метрики точности множественного отслеживания объектов (MOTA). Предложенный метод сравнивался с методом отслеживания на основе IoU и PCKh.

Основные результаты оценки алгоритма представлены в таблице 1. Как видно, данный подход к отслеживанию обеспечивает лучшие показатели MOTA и % IDSW, чем другие методы.

Исключение C_{center} или C_{pose} значительно увеличивает количество ошибок идентификации, указывая на необходимость обоих слагаемых.

Рис. 4 — Примеры правильно разрешенных окклюзий в видео. Разные цвета соответствуют разным ID.

В **четвертой главе** представлен новый стохастический метод решения задачи множественного отслеживания — Monte-Carlo Trajectory Optimization (MCTO). Под стохастическим методом понимается метод, использующий в своей работе генерацию псевдослучайных чисел. Для сравнения описаны два существующих стохастических подхода — Монте-Карло метод на марковских цепях (Markov chain Monte Carlo, MCMC) и объединенный вероятностный фильтр ассоциации данных Монте-Карло (JPDAF-MC). Ключевое отличие описываемого метода заключается в оптимизируемом пространстве. Вместо оптимизации связей между обнаруженными объектами и отслеживаемыми, и последующего уточнения траектории отслеживаемых объектов по полученным связям, предлагается оптимизировать траектории движения объектов, а затем строить связи с обнаружениями для оценки их правдоподобности.

Таблица 1 — Результаты верификации трекера PBBM на массиве данных PoseTrack 2017.

Tracking Algorithm	MOTA	%IDSW	Time, ms [a]
IoU-based	50.96	2.1	$\mathbf{0.11 \pm 0.18}$
IoU+PCKh-based	51.18	1.9	0.21 ± 0.54
Ours, w/o C_{center}, C_{vis}	50.39	2.6	0.21 ± 0.49
Ours, w/o C_{pose}, C_{vis}	50.91	2.1	0.16 ± 0.43
Ours, w/o C_{vis}	51.73	1.3	0.23 ± 0.52
Ours	**51.89**	**1.1**	5.14 ± 4.67

[a] Среднее время обработки трекером PBBM для одного кадра.

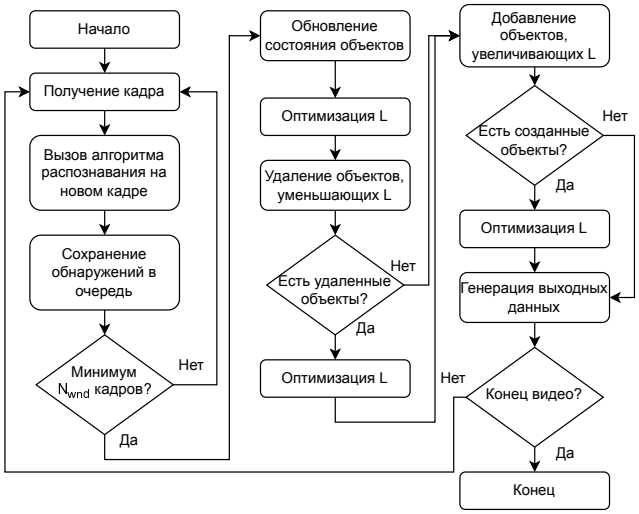

Рис. 5 — Блок-схема предложенного метода.

Блок - схема предложенного метода приведена на рисунке 5. Так как совокупное множество всех векторов $\vec{b}_{i,j}$ включающее все моменты времени и все объекты велико, то предлагается разбить весь временной интервал видео на пересекающиеся отрезки одинаковой длины — окна, и проводить оптимизацию в данных окнах независимо.

Первым шагом является получение входных данных для окна фиксированной длины в N_{wnd} кадров — координат обнаруженных объектов для каждого кадра с индексом из отрезка $[j, j + N_{wnd})$. После чего происходит поиск значений управляющих векторов $\vec{b}_{i,j}$ для каждого отслеживаемого объекта, максимизирующих оценку общей правдоподобности обнаружений $\mathcal{L}$. При этом $\vec{b}_{i,j}$ считается постоянным в данном окне.

Вероятные новые объекты создаются на основе обнаружений без ассоциаций в первом кадре окна. Для каждого из них происходит поиск оп-

тимальных значений управляющих векторов $\vec{b}_{i,j}$. Новый объект сохраняется, если его наличие увеличивает $\mathcal{L}$ на величину больше пороговой.

Затем удаляются объекты, отсутствие которых приводит к увеличению $\mathcal{L}$.

Оптимизация происходит отдельно для каждого окна:

$$\arg\max_{\vec{b}_{i,j} \in B} \mathcal{L}(\vec{b}_{0,j} \dots \vec{b}_{M-1,j} | \Theta) \tag{3}$$

где B — множество допустимых значений векторов $\vec{b}_{i,j}$, зависящее от специфики задачи, такой, как ограничения на величины ускорения и скорости, а функция правдоподобия $\mathcal{L}$ задает модель обнаружений.

Базовая структура функции $\mathcal{L}$ состоит из трех основных видов слагаемых — слагаемые, определяющие независимо вклад каждого трека, обнаружения и свойств самих векторов $\vec{b}_{i,j}$.

Для применения метода рассматривается следующая модель:

1. состояние объекта $\vec{a}_{i,j}$ включает в себя 4 компоненты - координаты p_x, p_y и проекции вектора скорости v_x, v_y.
2. управляющий вектор $\vec{b}_{i,j}$ состоит из двух компонент - проекций ожидаемой скорости в конце текущего окна u_x, u_y.
3. функция динамики задает изменение координат и скорости соответствующее равноускоренному движению временем Δt, с начальными координатами p_x, p_y, скоростью v_x, v_y и ускорением $\frac{\vec{u}-\vec{v}}{N_{wnd}\Delta T}$

$$\mathcal{L} = \sum_{t=0}^{N_{wnd}-1} C_t^{(wnd)} \Big(\sum_{i=0}^{M-1} F_{track}(|\Theta_t^{(i)}|) + \sum_{i=0}^{|\Theta_t|-1} F_{det}(\mathcal{T}, \Theta_{t,i}) \Big) + \sum_{i=0}^{M-1} F_{inertia}(\vec{b}_{i,j}) \tag{4}$$

Вклад трека F_{track} в величину $\mathcal{L}$ равен константе C_{miss} если отсутствуют ассоциированные обнаружения в предполагаемой точке существования объекта ($|\Theta_t^{(i)}| = 0$), и 0 в противном случае.

Вклад обнаружения F_{det} в величину $\mathcal{L}$ можно разбить на две составляющие — на оценку точности покрытия обнаружения и на штраф за наличие более чем одного ассоциированного трека. Точность покрытия оценивается по наиболее близкому изо всех ассоциированных объектов.

$$F_{det} = C_{hit} \max(1 - \frac{|\Delta x|}{\sigma}, 0) \times \max(1 - \frac{|\Delta y|}{\sigma}, 0) + C_{overlap} \max(0, |\Theta_t^{(i)}| - 1) \quad (5)$$

Здесь $\Delta x, \Delta y$ — разность координат обнаружения и координат ближайшего отслеживаемого объекта, C_{hit} — параметр алгоритма.

Второе слагаемое с коэффициентом $C_{overlap}$ вносит штраф за наличие более чем одного трека, перекрывающего обнаружение. Оно связано с гипотезой о том, что объекты не могут приближаться более чем на некоторое расстояние друг к другу.

Вклад свойств вектора $F_{inertia}$ в величину $\mathcal{L}$ соответствует априорной информации о распределении вектора $\vec{b}_{i,j}$. Была принята гипотеза о том, что движение с постоянной скоростью более вероятно, чем ускоренное — величина штрафа зависит от модуля разности скоростей в конце и начала окна:

$$F_{inertia} = C_{inertia} |\vec{u}_{i,j} - \vec{v}_{i,j}| \quad (6)$$

Оптимизируемая функция $\mathcal{L}$ не является выпуклой, и содержит в себе недифференцируемые слагаемые. Также весьма вероятно присутствие многих локальных минимумов. Поэтому было решено остановиться на рассмотрении стохастического метода оптимизации для поиска оптимальных значений $\vec{b}_{i,j}$, а именно метода локального случайного поиска. Причиной такого выбора является возможность значительно ускорить процесс вычисления $\mathcal{L}$ в случае если следующее высчитанное значение отличается от предыдущего незначительно.

Базовые значения параметров приведены в таблице 2. Если из описания эксперимента не следует иное, то используются значения из таблицы.

Обозначение	Описание	Значение	Диапазон
N_{wnd}	Размер окна	34	20–100
N_{iters}	Число итераций	150	20–192
C_{hit}	Вес попадания	1000	1000
C_{miss}	Вес пропуска	250	100–350
$C_{overlap}$	Штраф пересечения	150	150
N_{skip}	Пропуск кадров	1	1–7
$N_{iters}^{(add)}$	Число итераций при создании	250	250

Таблица 2 — Базовые значения используемых параметров.

Для валидации работы алгоритма и проверки качества отслеживания при использовании различных наборов параметров были использованы аннотированные видео из открытых наборов данных (датасетов) **MOT17, MOT20**. Метрика HOTA является рекомендуемой метрикой для сравнения результатов на MOT20.

В главе приводятся графики показывающие зависимость времени выполнения и качества отслеживания от основных параметров — размера окна и числа итераций оптимизации, а также возможные различные модификации алгоритма. Показано, что одним из основных и наиболее важных параметров является размер оптимизируемого окна в кадрах N_{wnd}. При его увеличении происходит уменьшение числа треков с увеличением их среднего качества, что полезно в применениях, где допустимо увеличение числа пропущенных объектов при высоком качестве траекторий. В терминологии precision и recall это соответствует увеличению precision при уменьшении recall.

Асимптотическая сложность алгоритма на кадр составляет $O(KNN_{wnd}P)$, где K - число итераций оптимизации, N - число обнаружений/объектов в кадре, N_{wnd} - размер окна, P - среднее количество объектов внутри одной ячейки разбиения. Проведенные эксперименты подтверждают линейную зависимость от числа итераций. Показано наличие квадратичной составляющей от среднего числа людей в кадре, если увеличение числа людей в кадре также связано с увеличением удельной плотности их количества на единицу площади кадра.

В **пятой главе** описывается реализация программного комплекса для построения траекторий людей на наблюдаемой территории, а также результаты прикладных применений. Приведены экспериментальные результаты, оценивающие требования к системе видеонаблюдения и необходимые вычислительные ресурсы.

Приводится описание компонент и общей структуры программного комплекса (рис. 6), включающего в себя ранее предложенные компоненты распознавания объектов, поз и реализацию двух новых методов отслеживания PBBM (глава 2) и MCTO (глава 3). Помимо этого, в главе описывается устройство компонент калибровки камеры, перевода СК, а также возможные конфигурации комплекса. Обязательной для выполнения компонентой

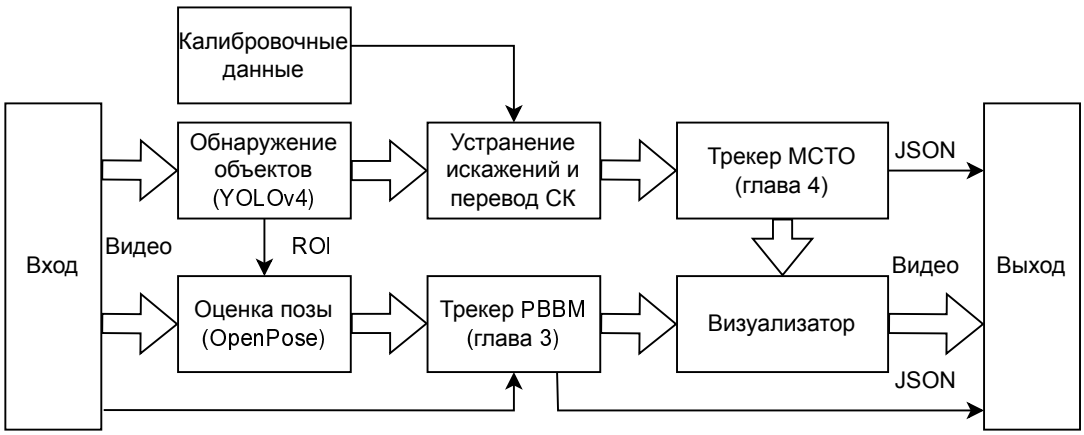

Рис. 6 — Схема программного комплекса.

является только компонента обнаружения объектов — все остальные могут свободно отключаться в зависимости от решаемой задачи.

Компонента перевода и устранения искажений решает задачу по переходу из двумерной системы координат пространства кадра, соответствующую пикселям на входном изображении (с началом в верхнем левом углу)в некоторую неподвижную трехмерную систему координат в реальном мире, в дальнейшем именуемую СК пространства сцены. Для этого проводится привязка обнаруженных объектов к некоторой точке в 3D пространстве сцены и определение координат этой точки с использованием данных о положении камеры. Это позволяет решить несколько задач: перейти к реальным физическим величинам вместо пикселей и тем самым упростить подбор параметров, точкам интереса на наблюдаемой территории и допустить совмещение измерений, выполненных несколькими камерами.

Общая идея перехода заключается в вычислении для каждого человека координат некоторой точки на изображении. После чего для каждой полученной точки рассчитываются параметры 3D луча, отображающегося в данную точку преобразованием камеры. По пересечению данного луча с известной поверхностью наблюдаемой 3D сцены можно найти 3D точку, соответствующую данному объекту. В программном комплексе реализовано несколько способов расчета исходной точки на изображении. Наиболее широко применимый способ заключается в использовании средней точки нижней грани ограничивающей рамки для объекта либо средней точки между координатами ступней человека.

Рассматриваются требования к вычислительным ресурсам для работы компоненты обнаружения объектов. Время обработки кадра изменяется почти линейно в диапазоне разрешений от 512 до 1024, и квадратично при учете больших разрешений. (график на рисунке 7) Линейная область графика обусловлена тем, что при малых размерах изображения при выполнении некоторых слоев нейросети загружены не все вычислительные ядра Nvidia RTX 2080 Ti. Подобные эффекты отсутствуют для памяти, и ее использование находится в достаточно строгой квадратичной зависимости от входного разрешения.

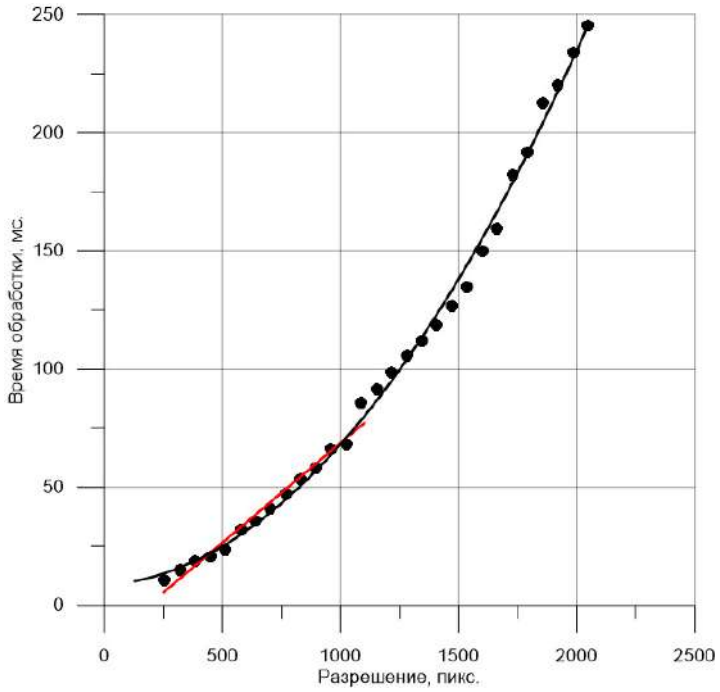

Рис. 7 — Время обработки одного квадратного кадра в зависимости от разрешения YOLOv4 на Nvidia RTX 2080 Ti.

В **шестой главе** приведены результаты практического применения разработанного программного комплекса, а также экспериментальные результаты, оценивающие требования к параметрам видео. В главе рассмотрено три прикладных применения — для решения задачи мониторинга очередей, для анализа спортивных матчей и для обнаружения нештатных ситуаций на эскалаторах.

В связи с тем, что работоспособность предложенных методов обнаружения и отслеживания может зависеть от ракурса камеры, нагруженности сцены, характерной скорости движения людей и других параметров, приво-

дится результат ряда вычислительных экспериментов отдельно для каждой прикладной задачи. Предлагается простая метрика для оценки результата распознавания:

$$Q = \frac{TP - FP}{N} \tag{7}$$

Где TP — число правильно распознанных объектов, FP — число ложных обнаружений, N - общее число объектов на изображении. Исследуется зависимость данной метрики от количества пикселей на отрезке, соответствующем высоте изображения человека для выбранного ракурса.

Задача мониторинга очередей была решена в рамках пилотного проекта выполненного АО "Центр открытых систем и высоких технологий"в Терминале В Московского Аэропорта Шереметьево. Основной целью проекта являлось формирование в реальном масштабе времени оценки наибольшего времени обслуживания человека в очереди. Было установлено несколько камер с различными ракурсами. (рис. 8)

При этом для решения задачи оценки математического ожидания времени обслуживания людей, стоящих в очереди, было необходимо решить следующие подзадачи:

1. обнаружить фигуры людей в кадре;
2. определить принадлежность людей к очереди и отфильтровать тех, кто проходит мимо или стоит вне связи с очередью;
3. определить и помнить уже прошедшее время ожидания каждого человека в очереди;
4. оценить оставшееся время ожидание человека до момента прохождения через точку обслуживания;
5. оценить максимальное время прохождения очереди пассажирами, стоящими в очереди.

Для решаемой прикладной задачи перечисленные задачи решаются с помощью разработанного программного комплекса.

В частности, определение принадлежности человека к очереди определялось на основе траектории его движения. Дальнейшая оценка времени ожидания формировалась методами, выходящими за рамки данной работы.

Особенностью задачи множественного отслеживания объектов для мониторинга очередей является относительно низкая скорость и высокая предсказуемость траекторий движения людей на наблюдаемой территории. При

этом характерны высокая плотность и большая степень корреляции пропусков обнаружений во времени. Корреляции связаны с тем, что изображение слабо изменяется если очередь стоит — и если сверточная нейронная сеть не в состоянии выделить объекты на некотором кадре, то и в последующих кадрах существует повышенная вероятность ошибки. С учетом данной специфики задачи, наиболее подходящим методом из предложенных является метод стохастического отслеживания.

В процессе проведения экспериментов были установлены факторы, влияющие на качество решения задач обнаружения объектов и построения траекторий — такие , как контрастный полосатый фон, наличие светового пятна, специфическое поведением людей в очереди при проходе на досмотр, заслоны от информационных стоек.

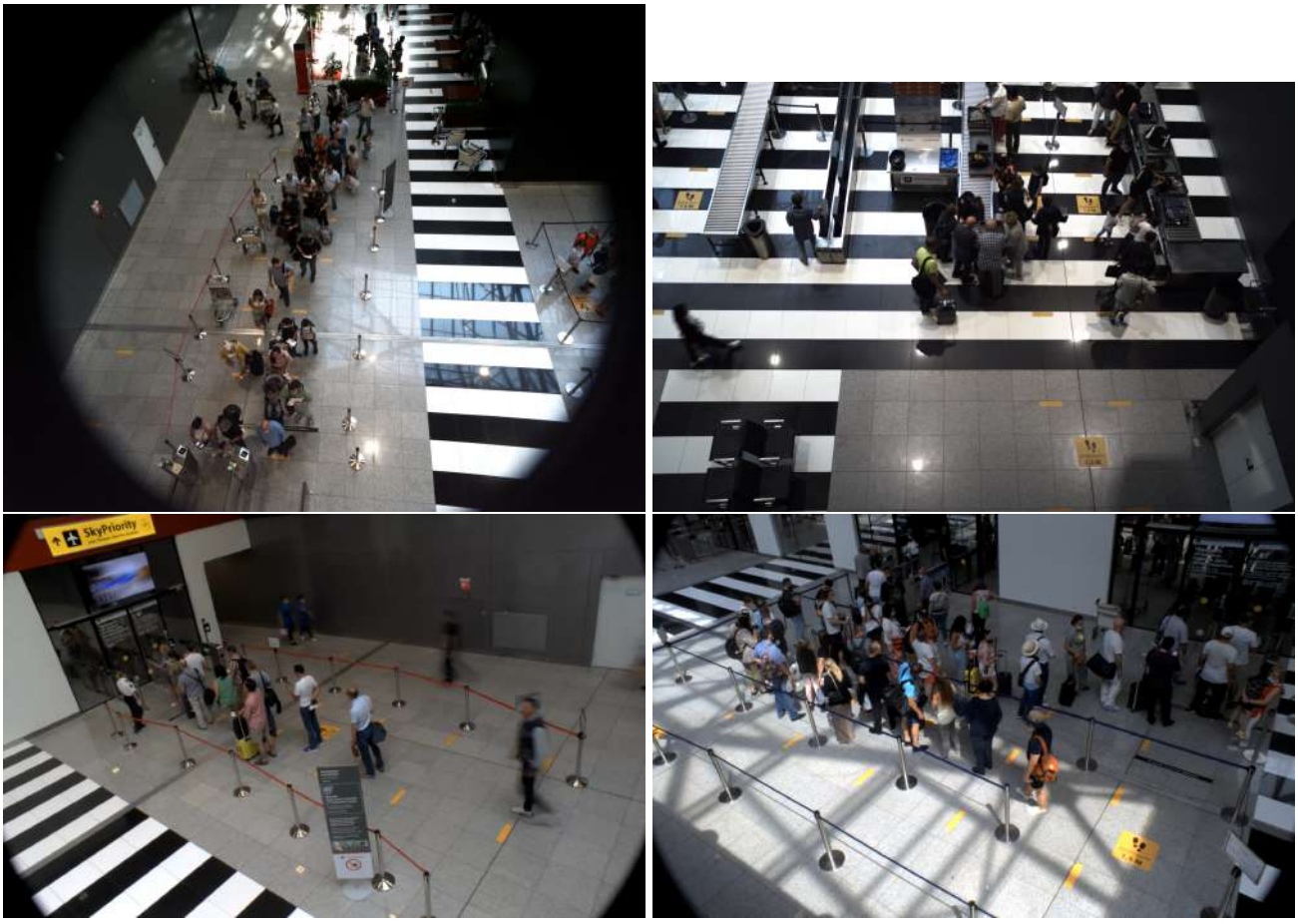

Рис. 8 — Очереди на обслуживание в Московском Аэропорту Шереметьево с различными ракурсами изображения.

Вторым рассматриваемым применением является их применение при анализе видеотрансляций игровых видов спорта. (рис. 9) Перед системой машинного зрения могут ставиться различные цели, такие как:

1. анализ видеотрансляции игры с целью предоставления тренеру дополнительной информации, упрощающей процесс тренировки;
2. анализ видеотрансляции игры с целью оценки эффективности игрока в реальном времени;
3. анализ видеотрансляции игры с целью оценки влияния конкретного игрока на игру;
4. анализ видеотрансляции игры для формирования аналитики и ее вывода в виде дополнительной информации , включаемой в видеотрансляцию.

Для проведения экспериментов были использованы запись 4K формата матчей испанской Премьер лиги (La Liga) по футболу, чемпионата мира 2020 года по бадминтону, а также собственные записи высшей лиги по гандболу России среди молодежных команд.

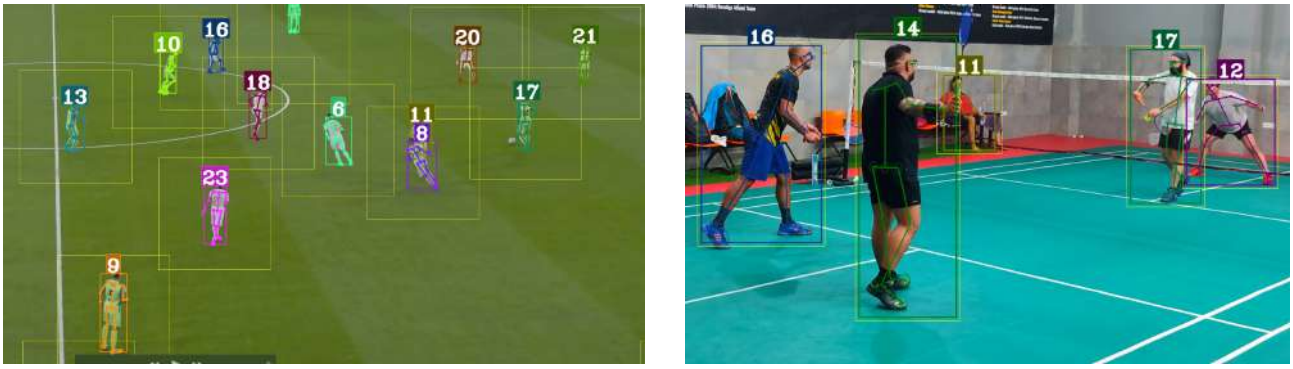

Рис. 9 — Фрагмент изображения трансляций игр в футбол и в бадминтон с обнаруженными фигурами игроков

Отмечается разница в требованиях к разрешению для успешного распознавания, по сравнению с применением для анализа очередей . Улучшения качества распознавания не было замечено начиная со средней высоты человека в ~ 100 пикселей, против ~ 160 для анализа очередей. Этот факт можно связать, в первую очередь, с ярко выраженным контрастом игроков в спортивной форме на фоне поля, а также с отсутствием дополнительных объектов на поле, усложняющих распознавание. (таких как сумки и багаж) Также для движения спортивных игроков характерно наличие больших и часто изменяющихся ускорений при активном передвижении по площадке, что ухудшает качество приближения траектории в окне траекторией с постоянным ускорением.

Поэтому предпочтительным является выбор трекера PBBM для данного применения.Применение компоненты распознавания позы позволяет значительно упростить процедуру построения алгоритма распознавания ТТД, как на основе машинного обучения, так и на основе эвристических алгоритмов.

Еще одним возможным применением разработанных решений является использование системы машинного зрения для обнаружения инцидентов на эскалаторах в реальном времени. Под инцидентами службы аэропорта понимают события, ставящие под угрозу безопасность пассажиров — падения людей, багажа, передвижение в положении сидя, и т. д.

Для апробации метода был установлен набор камер, осуществляющих видеорегистрацию на эскалаторах и пассажирских конвейерах в Московском Аэропорту Шереметьево (рис. 10).

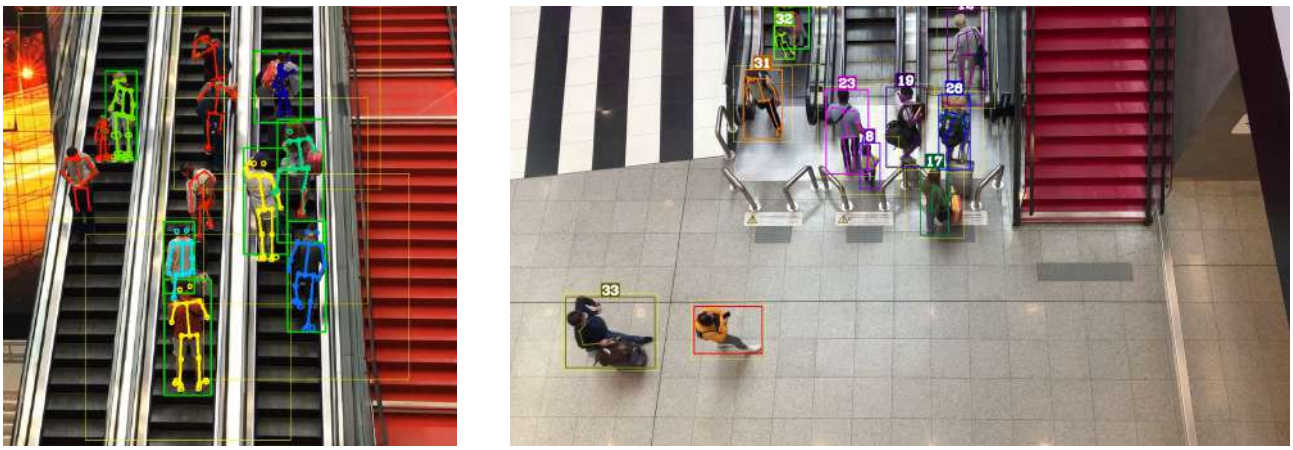

Рис. 10 — Примеры фрагментов кадров для обнаружения инцидентов на эскалаторах

В **заключении** приведены основные результаты работы, которые заключаются в следующем:

1. Приведен обзор нейросетевых методов локализации объектов на изображениях и методов множественного отслеживания объектов. Проанализирована применимость различных методов для решения задачи сопровождения людей в высоконагруженных сценах.
2. Предложен, программно реализован и апробирован метод сопровождения людей, представленных скелетной моделью, на основе двудольного сопоставления (PBBM – Pose-based Bipartite Matching).

3. Работоспособность данного метода проверена на открытых наборах данных PoseTrack17, PoseTrack18.
4. Предложен, программно реализован и апробирован стохастический метод сопровождения объектов с оптимизацией в скользящем окне (MCTO – Monte-Carlo Trajectory Optimization).
5. Проведено множество численных экспериментов, показывающих производительность данного метода в зависимости от параметров и оценивающих качество отслеживания на открытых размеченных наборах данных MOT17, MOT20.
6. Для выполнения поставленных задач разработан программный комплекс, комбинирующей реализацию вышеперечисленных методов и открытые реализации нейросетей OpenPose, YOLO, а также ряд вспомогательных алгоритмов по переводу систем координат и устранению оптических искажений.
7. Разработанный программный комплекс успешно использован при решении реальных задач по мониторингу очередей в АО “Московский аэропорт Шереметьево”.
8. Разработанный программный комплекс успешно использован при решении реальных задач по анализу видео в игровых видах спорта.

Результаты работы апробированы на международных научно - технических конференциях:

1. 4th International Conference on Electrical, Control and Instrumentation engineering (ICECIE), Kuala-Lumpur, Malaysia, 2022
2. Международной конференции Computing Conference 2021, Лондон, Великобритания, 15 июля 2021;
3. XXII Международной конференции «Цифровая обработка сигналов и ее применение DSPA-2020» ИПУ им.Трапезникова";
4. Международной конференция Intelligent Systems Conference (IntelliSys 2020), май 2020, Амстердам, Нидерланды;
5. Международной конференции International Conference on Technology and Entrepreneurship (ICTE), Болонья, Италия, 20-21 апреля 2020

Работа выполнена при поддержке гранта РФФИ № 19-29-09090 “Разработка методов и технологий анализа видеоинформации в распределенных ге-

терогенных системах видеонаблюдения с использованием дискретной модели наблюдаемой сцены и информации о пространственно - временной привязке видеопотоков и файлов”

Список публикаций

1. Gilya-Zetinov A., Bugaev A., Khelvas A., Konyagin E., Segre J. High-Speed Multi-person Tracking Method Using Bipartite Matching // Lecture Notes in Networks and Systems, 2022, 283, pp. 793–809

2. Zuev I., Gilya-Zetinov A., Khelvas A., Konyagin E., Segre J. Humans Digital Avatar Reconstruction for Tactical Situations Animation //Lecture Notes in Networks and Systems, 2022, 283, pp. 634–644

3. Khelvas A., Gilya-Zetinov A., Konyagin E., Demyanova D., Sorokin P., Khafizov R. Improved 2D Human Pose Tracking Using Optical Flow Analysis // Advances in Intelligent Systems and Computing, 2021, 1251 AISC, pp. 10–22

4. Khelvas A., Demyanova D., Gilya-Zetinov A., Konyagin E., Khafizov R., Pashkov R. Adaptive distributed video surveillance system // 2020 International Conference on Technology and Entrepreneurship - Virtual, ICTE-V 2020, 9113774

5. Хельвас А.В., Хафизов Р.Р., Гиля-Зетинов А.А., Малышев С.А. Разработка архитектуры программной AI платформы для анализа тактико-технических действий и функционального состояния футболистов в процессе игры по данным видеотрансляции // Доклады на 22-ой Международной конференции. Сер. "Цифровая обработка сигналов и её применение"Москва, 2020, Российское научно-техническое общество радиотехники, электроники и связи им. А.С. Попова, стр. 652-656

6. Monte-Carlo based 2D object tracking approach in high load scenes / A.Gilya-Zetinov, E. Tsybulko, A.Bugaev, A. Khelvas, A. Zaitseva, R. Pashkov // Proceedings of 2022 4th International Conference on Electrical, Control and Instrumentation Engineering (ICECIE). — IEEE. 2022. — С. 102—112. — in publishing