

	<p>Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»</p>	Лист 1/17
--	---	-----------

Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности

Документация

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 2/17
--	---	-----------

ОГЛАВЛЕНИЕ

1. Термины, определения и сокращения	3
2. Общие положения	4
2.1. Описание системы	4
2.2. Назначение документа	5
3. Требования к системе	5
3.1. Функциональные требования	5
3.2. Нефункциональные требования	6
3.3. Виды нарушений	6
3.4. Ограничения	7
4. Техническое решение	7
4.1. Общее описание и принцип работы системы	7
4.2. Перечень программного обеспечения	8
4.3. Ограничения системы	9
4.4. Архитектура	9
4.5. Логика обнаружения нарушений	15
4.6. Пользовательский интерфейс	15
4.7. Безопасность	16
4.8. Лицензирование	17

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 3/17
--	---	-----------

1. Термины, определения и сокращения

В настоящем документе используются следующие термины, определения и сокращения, представленные в Таблица 1.

Таблица 1 – Термины, определения и сокращения

Термин	Описание
СИЗ	Средства Индивидуальной Защиты
RTSP	Real-Time Streaming Protocol, протокол для передачи видеопотоков в режиме реального времени.
FPS	Frame Per Second, количество кадров в секунду.
SoftAccuracy	$SoftAccuracy = \frac{1}{2} \left(\frac{T}{S} + \left(1 - \alpha \cdot \frac{F}{N-S} \right) \right) \cdot 100\%$ <p>где: значение softAccuracy изменяется в диапазоне от 0 до 100% (чем больше, тем лучше), включительно, S - заранее, априорно известное (GT) число видео последовательностей с нарушениями правил ношения СИЗов; N – априорно известное (GT) общее количество видео последовательностей на тестовой выборке, длина каждого порядка t секунд. (требование: N>S); T - количество видео последовательностей с корректным срабатыванием разработанной системы (штатная работа система, в противоположность GT); F – количество видео последовательностей с ложным срабатыванием разработанной системы (штатная работа система, в противоположность GT); $\alpha > 1$ - регуляризационный параметр ошибки 2-го рода.</p>
UML	Unified Modeling Language, язык графического описания для объектного моделирования в области разработки программного обеспечения, для моделирования бизнес-процессов, системного проектирования и отображения организационных структур.
BPMN	Business Process Model and Notation, система условных обозначений и их описания в XML для моделирования бизнес-процессов.
YOLO	Модель машинного обучения для детекции объектов и сегментации (You Only Look Once).

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 4/17
--	--	-----------

Термин	Описание
Трекинг	Процесс отслеживания перемещения объектов между последовательными кадрами.
Kafka	Распределенная стриминговая платформа для обработки потоковых данных.
KsqlDB	Модуль потоковой обработки данных на основе прописанных SQL подобных запросов, адаптированный для работы с Kafka
MinIO	Высокопроизводительное хранилище файлов.
Postgres	PostgreSQL, реляционная база данных с открытым исходным кодом.
InSight	Корпоративная автоматизированная система управления процессами промышленной безопасности и охраны труда.
REST	Архитектурный стиль взаимодействия компонентов распределённого приложения в сети.
API	Application Programming Interface, интерфейс программирования приложений.
ML	Machine Learning (машинное обучение).
JWT токен	JSON Web Token, ключ аутентификации пользователя.

2. Общие положения

2.1. Описание системы

Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности **предназначена** для выявления случаев нарушения промышленной безопасности, неправильного ношения СИЗ (средств индивидуальной защиты) работниками предприятия, нахождения в запрещенных зонах с помощью камер видеонаблюдения в режиме реального времени с целью оперативного реагирования на нарушение техники безопасности. Система подразумевает возможность расширения путем добавления новых детекторов, позволяющих распознавать новые типы событий.

Система позволяет:

- сократить случаи травматизма работников и несчастных случаев вследствие неприменения средств индивидуальной защиты (СИЗ) и нарушения техники безопасности;
- обеспечить своевременное оповещение ответственных лиц об опасностях и нестандартных ситуациях.

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 5/17
--	---	-----------

Входные и выходные данные:

Система принимает на вход данные с камер видеонаблюдения, установленных на предприятиях и в режиме реального времени детектирует различные нарушения формируя периодическую отчетность по ним, которую можно просматривать в веб-интерфейсе системы, скачивать отчеты в различных форматах. Предусмотрены уведомления посредством передачи email и sms (опционально) сообщений.

Функциональные характеристики программного обеспечения

- Система поддерживает входные видеопотоки с веб камер, из файлов, с IP камер
- Система включает модуль распознавания лиц для определения нарушителя
- Система способна работать с моделями классификации, детекции и сегментации
- Система способна рисовать ограничивающие рамки для задач детекции, маски для задач сегментации и писать метки для задач классификации, в том числе текстовые
- Система поддерживает несколько источников ввода
- Система способна работать с различными архитектурами нейронных сетей
- Система имеет модуль трекинга для подключения к моделям детекции
- Система поддерживает как с CPU, GPU так и TPU
- Система позволяет сохранять веса моделей машинного обучения и их архитектуру
- Система включает модуль разметки и дообучения для включения новых событий и объектов (новых типов СИЗ, зон, предметов и т.д.)

2.2. Назначение документа

Документ описывает программное обеспечение Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности (далее — Системы).

Документация разработана в соответствии с техническим заданием, ГОСТ 19.105-78, ГОСТ 19.102-80.

3. Требования к системе

3.1. Функциональные требования

Система удовлетворяет следующим функциональным требованиям.

- **Видеоаналитика:** система считывает данные с предоставляемых видеопотоков, определяет человека и наличие/отсутствие требуемых СИЗ на кадре и выявляет нарушения правил ношения СИЗ, проникновение в запрещенные места и зоны. Перечень детектируемых СИЗ может расширяться путем добавления новых детекторов.

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 6/17
--	--	-----------

- **Формирование и предоставление отчетности:** система обеспечивает формирование автоматизированной отчетности, имеет возможность предоставлять отчетную информацию в табличном и графическом виде.

- **Интеграция со сторонними сервисами:** система позволяет интегрироваться со внутренними системами предприятия посредством API (программного интерфейса взаимодействия).

- **Управление доступом:** система должна иметь функционал по управлению доступом пользователей к системе в рамках модели разграничения доступа.

3.2. Нефункциональные требования

Система удовлетворяет следующим нефункциональным требованиям.

- **Разграничение доступа:** доступ к определённым функциям системы осуществляется в рамках ролевой модели разграничения доступа.

- **Масштабируемость:** система допускает увеличение количества одновременно анализируемых видеопотоков с сохранением работоспособности в рамках ограничений серверной инфраструктуры.

- **Качество: детектирование** человека на кадре и определение наличия или отсутствия на рабочем уставных элементов СИЗ производится на расстоянии в диапазоне 3 ... 12 метров от камер с точностью не хуже SoftAccuracy ~70%.

- **Хранение:** хранение информации о нарушениях ношения СИЗ производится в течение одного года, но может быть отдельно настроено.

- **Резервное копирование:** резервное копирование и восстановление осуществляется силами и средствами пользователя.

- **Формат событий нарушения:** представляет информацию о параметрах СИЗ, дате, времени и локации нарушения.

- **Безопасность:** система использует безопасные протоколы обмена информации при межмодульном взаимодействии и при интеграции с другими системами.

- **Патентная чистота:** компоненты системы не требуют дополнительных лицензий на программное обеспечение сторонних производителей.

3.3. Виды нарушений

Перечень средств индивидуальной защиты, встроенных в систему по умолчанию представлен в Таблица 2.

Таблица 2 – Определяемые виды СИЗ

Наименование СИЗ	Расположение СИЗ
Защитная каска	На голове работника
Подбородочный ремень	На подбородке работника

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 7/17
--	---	-----------

Защитные очки	На голове работника
Фонарь	На голове работника
Костюм от ОПЗ, застегнутый на все пуговицы	На теле работника
Самоспасатель	Через плечо работника

3.4. Ограничения

- **Количество камер:** неограниченное количество, в зависимости от возможностей оборудования.

- **Условия среды:** контроль с камер производится в закрытом помещении круглосуточно без факторов воздействия окружающей погодной/сезонной среды.

- **Освещённость:** освещённость целевой площадки обеспечивается не менее 50 лк (согласно российским нормам по СНиП 23-05-95).

4. Техническое решение

4.1. Общее описание и принцип работы системы

Система видеоаналитики представляет собой многокомпонентную архитектуру.

Видео с камер с помощью RTSP протокола передаются в систему считывания видеопотока, которая использует аппаратные возможности системы для декодирования видеопотока и разбиения его на отдельные кадры. Полученные кадры вместе с метаданной о камере попадают в конвейер данных, представляющий из себя сервер Kafka. Дополнительно данные кадры сохраняются во временное хранилище MinIO для минимизации затрат на передачу их по всему конвейеру Kafka.

По конвейеру данных кадры с камер попадают в ML обработчики, представляющие из себя модули по обнаружению СИЗ, людей и их ключевых точек, трекинг найденных людей.

Дальше с помощью конвейера данные о найденных людях и СИЗ передаются в модуль обнаружения нарушений. В модуле происходит вычисление целевого места ношения СИЗ и сопоставление его с обнаруженными СИЗ. Это необходимо для выявления присутствия/отсутствия СИЗ у человека. На основании данной информации формируется вероятность нарушения заданных правил ношения в рамках одного конкретного кадра для каждого человека.

После обнаружения нарушения по конвейеру данных попадают в модуль агрегации нарушений во времени для сбора статистики нарушений за весь период присутствия человека в кадре. После окончания агрегации, когда человек покидает кадр, собранные статистические данные попадают в модуль выявления нарушений. Данный модуль отвечает

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 8/17
--	---	-----------

за вынесение финального решения о нарушении человеком правил ношения СИЗ на основе превышения суммарной уверенности в нарушении определённого порога.

Взаимодействие пользователей с системой производится с помощью веб-интерфейса, который в свою очередь общается с системой с помощью REST API. В рамках взаимодействия пользователю предоставляется возможность по:

- управлению учётными записями пользователей;
- управлению камерами, используемыми системой;
- просмотр информации о найденных нарушениях.

4.2. Перечень программного обеспечения

Перечень устанавливаемого системного и прикладного ПО на серверах Системы приведен в Таблица 3.

Таблица 3 – Перечень программного обеспечения

Термин	Описание	Технология
ОС сервера	Операционная система для работы серверов	Ubuntu Server 24.04
YOLO	Модель машинного обучения для детекции объектов	YOLOv8
Apache Kafka	Платформа для передачи сообщений и потоков данных	Apache Kafka
KsqlDB	Надёжная система агрегации нарушений	KsqlDB
MinIO	Временное хранилище изображений с камер	MinIO
PostgreSQL	Система управления базами данных для хранения данных о нарушениях	PostgreSQL 16
Nvidia CUDA	Библиотека для обработки данных на GPU	CUDA 12.x
Docker	Контейнеризация компонентов системы	Docker 20.x
Python	Язык программирования для серверной части системы	Python 3.11

Система использует контейнеризацию на базе Docker, что упрощает развертывание и масштабирование компонентов. Использование моделей YOLO требует оптимизации под GPU, что достигается при помощи Nvidia CUDA. Apache Kafka позволяет масштабировать обработку потоков, а PostgreSQL обеспечивает долговременное и надежное хранение данных.

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 9/17
--	---	-----------

4.3. Ограничения системы

Для обеспечения корректного поведения системы на входящий видеопоток накладываются следующие ограничения.

- **Цвет:** должен быть трёхканальным в RGB пространстве.
- **Качество изображения:** должно позволять различать целевые СИЗ с расстояния как минимум 12 метров от камеры.
- **Близость людей:** целевой поток людей, на которых планируется обнаружение нарушений, должен проходить не дальше 12 метров от камеры для обеспечения возможность различать все рассматриваемые типы СИЗ.
- **FPS:** должно быть не ниже 10 кадров/сек.

Также, ввиду подхода по минимизации ложный срабатываний, нарушения формируются только спустя несколько минут после выхода человека из области камеры.

4.4. Архитектура

Основными модулями системы являются следующие.

Считыватель видеопотока

Производит считывание кадров из RTSP потока, производит пропуск кадров до требуемого FPS, обрезает целевую область камеры. Подготовленное изображение отправляет во временное хранилище и, вместе с информацией о камере, в конвейер данных Kafka.

Конвейер данных

Для реализации непрерывной потоковой обработки данных с видеопотоков была выбрана Kafka. Kafka является надёжной системой по управлению очередями сообщений, получению и распределению данных по связанным системам.

Политика очистки – спустя 3 часа записанные данные стираются для экономии памяти системы. Этого времени достаточно, чтобы модули обработали хранящуюся информацию, даже если возникнет временный сбой или перезапуск некоторых компонентов.

Временное хранилище кадров

Для реализации временного хранилище кадров была использована MinIO. MinIO была выбрана как удобный и надёжный сервер по хранению файлов изображений. MinIO используется как альтернатива чтению конкретных сообщений из очереди Kafka, которое в ходе испытаний показало свою нестабильность.

Политика очистки – спустя 3 часа записанные данные стираются для экономии памяти системы. Этого времени достаточно, чтобы абсолютное большинство пользователей смогло покинуть область камеры для завершения агрегации с последующим сохранением кадра в постоянное хранилище.

ML обработчики

Производит детектирование СИЗ, людей и их ключевых точек с помощью обученных на предоставленном заказчиком наборе данных моделях YOLO. Также отвечает

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 10/17
--	---	------------

за трекинг людей между входной последовательностью кадров с помощью алгоритма ВоТ-SORT для обеспечения возможности последующей агрегации данных во времени.

Каждая группа обработчиков подключена только к одному видеопотоку для возможности масштабирования и обеспечения условий работы алгоритма трекинга.

Результатом обработки являются: ограничивающие рамки и уверенность детектирования СИЗ, ограничивающие рамки, ключевые точки конечностей и уверенность детектирования всех людей на рассматриваемом кадре вместе с присвоенными идентификаторами людей. Данная информация объединяется для последующего анализа на наличие нарушений.

Детектор нарушения СИЗ

На основании данных от ML обработчиков производит поиск целевых областей СИЗ, сопоставление найденных объектов СИЗ с человеком, сопоставление объектов СИЗ человека с целевыми областями.

Поиск целевых областей СИЗ происходит на основании положений ключевых точек конечностей человека с помощью прописанной логики для каждого типа СИЗ. Формирование данной логики производилось эмпирически и отражает расположение СИЗ, описанное в Таблица 2.

Сопоставление объектов происходит с помощью расчёта площади пересечения ограничивающих рамок относительно площади их объединения. При превышении установленного порога два объекта считаются сопоставленными. Если для одного объекта СИЗ происходит сопоставление сразу с несколькими людьми, то сопоставление происходит на основании близости к целевой области СИЗ для каждого человека в отдельности. Один объект СИЗ может быть сопоставлен только с одним человеком.

На основании сопоставления объектов СИЗ человека с целевыми областями выносится решение об отсутствии соответствующего типа СИЗ (нарушении) у каждого человека с определённой уверенностью на рассматриваемом кадре. Если сопоставление прошло успешно, то уверенность равна 0, иначе уверенность зависит от уверенности детектирования человека и объекта СИЗ.

Агрегатор нарушений

Производит агрегацию нарушений во времени для каждого нарушения и человека, вычисляя среднюю уверенность нарушения и человека, дату и время наилучшего кадра. Наилучшим кадром считается тот, высота человека на котором максимальна (чем ближе к камере – тем лучше виден).

Для уменьшения количества записываемых нарушений дополнительно объединяет вычисленные статистики нарушений по дате и времени кадра по каждому человеку.

Для реализации агрегации используется система KsqlDB.

Модуль формирования нарушений

Производит окончательное формирование нарушений на основе собранных статистик модулем агрегации нарушений. При превышении порога список нарушений с информацией о человеке и исходном кадре отправляются дальше для их фиксации.

Архиватор нарушений

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 11/17
--	--	------------

Производит получение нарушений из конвейера данных; получение исходных кадров, на которых эти нарушения были зафиксированы; сохранение информации о нарушениях в хранилище.

REST API

Предназначен для предоставления конечных точек доступа к функциям системы посредством взаимодействия через HTTP протокол согласно политике REST. Является промежуточным звеном между пользовательским интерфейсом и системой и реализует внутри себя логику данного взаимодействия.

Основой сервиса является библиотека FastAPI.

Серверное оборудование

При использовании камер (видеопотоков 30 fps) в количестве 1-2 шт., при условии достаточности памяти, система разворачивается на одном физическом сервере с одной GPU. Уменьшение количества обрабатываемых кадров в секунду (fps) и увеличение количества GPU пропорционально увеличивает количество поддерживаемых камер.

При увеличении количества камер, ядро системы разворачивается на нескольких серверах, а обмен данным происходит через распределенную шину данных (kafka).

Дальнейшее расширение системы (в рамках предприятия) достигается путем горизонтального масштабирования и добавления серверов

По итогам анализа отчетов производительности выявлено, что RTX 4090 обрабатывает от 300 до 500 детекций (инференсов) в секунду на моделях YOLOv8m (medium) и 200-250 на моделях YOLOv8x (xlarge, 1080p) соответственно. Система использует минимум 5 моделей при анализе каждого кадра.

При обработке каждого кадра (25-30fps) масштабирование системы рассчитывается следующим образом: 1 GPU RTX 4090 для 5 новых камер. При этом в соответствии с результатами тестирования системы выявлено, что для устойчивой детекции СИЗ достаточно 5fps. Таким образом (при применении 5 fps), для модели размера YOLOv8x максимальная пропускная способность составляет: 8-10 камер / GPU, а для YOLOv8m пропускная способность 12-25 камер / GPU соответственно, что позволит существенно сократить расходы на оборудование.

Рекомендуется учесть тот факт, что при увеличении количества видеопотоков, требуется объем не менее 4Gb RAM непосредственно на видеопоток, в том числе на сопутствующие ему kafka-топики, веса ML-моделей.

Минимальная инфраструктура. 1-2 камеры, 1 сервер

Прогнозные требования:

- Обработка 2 видеопотоков в 1080p при 30 кадрах в секунду.
- Обработка каждого кадра с помощью 5 разных моделей машинного обучения (YOLO).
- Хранение видеопотоков в Kafka с retention policy в 8 часов.
- Долговременное хранение только 100-200 кадров в сутки.

Процессор:

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 12/17
--	---	------------

Для параллельной обработки потоков и управления системами ввода/вывода (Kafka, Postgres, модельный сервер) сервер должен быть оснащён мощным многоядерным процессором:

- Процессор: AMD Ryzen Threadripper PRO 5955WX или аналоги
- Количество ядер/потоков: 16 ядер / 32 потока.
- Базовая частота: 4.0 ГГц (с бустом до 4.5 ГГц).

Преимущества: Высокая производительность на одно ядро и возможность обрабатывать множество параллельных задач. Threadripper обладает характеристиками для обработки ввода/вывода (что важно для Kafka), а также для управления GPU-процессами и управления системными ресурсами.

16 ядер является достаточным показателем для управления двух видеопотоков и всей сопутствующей системы, при этом оставляя аппаратные мощности для администрирования, передачи данных и выполнения вторичных задач (мониторинга, логирования).

Графические карты (GPU):

Для обработки видеопотоков с YOLO необходимо использовать графические процессоры для распознавания объектов в режиме реального времени.

- Графическая карта: Nvidia RTX 4090/4070/A4000/A5000 (2 шт.)
- Количество CUDA-ядер: 16 384 (RTX 4090).
- Память: 24 ГБ GDDR6X (RTX 4090).
- TFLOPS: ~82.6 TFLOPS FP32 производительности.

Обоснование расчёта:

Видеопотоки: 2 видеопотока в 1080p, 30 FPS, требуют 60 кадров в секунду.

ML-Модели: в качестве примера приводятся следующие расчеты применимо к моделям YOLOv8x. На каждый кадр требуется запустить 5 моделей. Таким образом, для обработки $60 \text{ к\c} \times 5$ (кол-во видеокамер) = 300 inferences в секунду.

Оценка производительности: на одной RTX 4090 каждая модель YOLOv8x в разрешении 1080p обрабатывает до 200-250 inferences в реальном времени (в зависимости от модели), при этом удерживая 50-60 FPS.

Для оптимизации используемых ресурсов имеется возможность видеопотоки с двух камер наблюдения направить на одну видеокарту, однако, чтобы иметь резерв на дополнительные модели, дообучение и другие процессы, рекомендуется использовать две видеокарты.

Две видеокарты NVIDIA GeForce RTX 4090 обеспечивают стабильную работу с запасом мощности для обработки 300 детекций в секунду.

Оперативная память (RAM):

Для корректной работы системы Kafka, Postgres и различных сервисов (уведомления, администрирование и т.д.), а также для кэширования данных и работы с кадрами, требуется обеспечить достаточный объем оперативной памяти.

Оперативная память: 128 GB DDR4 ECC.

Обоснование:

Kafka: Поточковые данные занимают несколько десятков гигабайт при 8-часовом периоде хранения. Память будет использоваться для кэширования данных в Kafka и для работы всех сервисов в системе.

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 13/17
--	---	------------

Модели ML: для запуска каждой модели требуются несколько гигабайт памяти на временные данные и управление процессами обработки.

Хранилище (SSD):

Хранилище гарантированно обеспечивает высокую скорость ввода-вывода для работы с видеорядом и кадрами.

Обоснование: системы, работающие с видеоданными, требуют высокой скорости доступа к файлам и временным данным. NVMe обеспечит низкие задержки и высокую скорость записи/чтения, что важно для Kafka, а также для хранения временных кадров в minio и реляционных данных в Postgres.

Основное хранилище (SSD):

Один диск **1 ТБ NVMe SSD** для ОС и телеметрии,

Другой диск **2 ТБ NVMe SSD** для Kafka и Postgres.

Kafka хранит кадры из двух видеопотоков за последние 8 часов:

- Видео 1080p 30 FPS генерирует примерно 5 Мбит/с данных.
- Объем данных составляет порядка $5 \text{ Мбит/с} * (8 * 3600) \text{с} * 8 / 1024 \sim 18 \text{ ГБ}$ данных в период 8 часов на 1 видеопоток.
- Для 2 видеопотоков: $18 \text{ ГБ} * 2 = 36 \text{ ГБ}$ данных двух очередей с кадрами в Kafka.
- С учетом реализации кадры проходят через несколько очередей внутри системы, тем самым временно формируя дублирование для разных компонентов.
- Требуется учесть дополнительно репликацию самой Kafka как минимум x2 (т.к. replication factor = 2 — означает, что необходимо хранить 2 копии данных)
- С учётом небольшого объёма метаданных, около 500 ГБ SSD достаточно для Kafka, Postgres Pro и системных процессов.
- Регулярная перезапись одних и тех же участков SSD выводит накопитель за короткий промежуток времени из строя. Предлагается использовать SSD большего объема, контроллер распределит операции записи более равномерно, тем самым продлевая срок службы SSD накопителя.
- Требуется учесть дополнительно возможное расширение системы и увеличение числа камер до 3-4 (если позволит GPU).

Распределенная инфраструктура. (10-20 камер, 2-4 GPU-сервера)

Основные параметры системы:

- Количество камер: 10-20 видеопотоков с разрешением 1080p, 30 FPS.
- Обработка кадров: Каждый кадр обрабатывается 5 моделями YOLO.
- Retention в Kafka: 8 часов хранения видео в промежуточном буфере Kafka.
- Долговременное хранилище: 100-200 кадров в сутки.

Основные компоненты инфраструктуры:

- Видеопроецессинг: Основная нагрузка по инференсу распределяется на GPU.
- Поточковая передача данных (Kafka): Видео передается через Kafka, что позволяет распределять нагрузку и временно хранить данные.
- Долговременное хранилище: Хранение кадров на дисковом хранилище.

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 14/17
--	---	------------

- Сервера (CPU, GPU, RAM, Storage): Основной расчет идет на использование стандартных стоечных серверов с достаточным числом ядер CPU и мощными видеокартами.
- Кластер GPU-серверов:
- 2-4 сервера с GPU видеокартами, такими как Nvidia 4090 или модели более позднего ряда, например NVIDIA Tesla A100
- Каждая машина может обрабатывать до 5 видеопотоков на 1 видеокарту RTX 4090, при разрешении видеопотока 1080p на YOLOv8 с высокой частотой кадров (~30 FPS).
- Загрузка GPU на одной машине может составлять 80-90%
- Kafka для обработки данных и балансировки:
- Используется для передачи потоков видеокладов от RTSP к серверам с GPU для обработки.
- Kafka имеет функцию распределения задания между несколькими серверами с GPU для балансировки нагрузки.
- Postgres Pro + minio:
 - Метаданные нарушений и отчёты продолжают храниться в Postgres Pro
 - Видео или отдельные кадры сохраняются на внешних серверах хранения, предположительно в minio.

Рекомендуемая конфигурация:

Требуется кластер из 4 серверов. Каждый сервер будет оборудован для обработки 10 видеопотоков в режиме 30 fps. Конфигурация серверов включает в себя высокопроизводительные процессоры для управления системными процессами и видеокарты для инференсов.

Конфигурация одного сервера:

- Процессор (CPU): AMD Ryzen Threadripper PRO 5955WX или аналог.
- 16 ядер / 32 потока.
- Частота: 4.0-4.5 ГГц.

Обоснование: Threadripper обладает высокой производительностью для параллельных задач, таких как обработка потоков видео, управление Kafka, Postgres Pro, системными сервисами.

- Графическая карта (GPU): 2 x Nvidia RTX 4090 (или аналог: A4000/A5000).
- Количество CUDA-ядер: 16 384 на одну видеокарту.
- TFLOPS FP32: 82.6 TFLOPS на одну видеокарту.

Обоснование: для обработки видеопотоков от 5 до 10 камер в реальном времени с 5 моделями YOLO требуется мощное графическое оборудование. Две RTX 4090 на один сервер обеспечат достаточную вычислительную мощность для обработки инференсов по 10 потокам с запасом.

Оперативная память (RAM): 256 GB DDR4 ECC.

Обоснование: Kafka, Postgres Pro и сами ML-модели требуют значительных объемов оперативной памяти для временного хранения данных. С учетом параллельной обработки потоков и кэширования данных в памяти, 256 ГБ на сервер обеспечат стабильную работу системы.

Хранилище (NVMe SSD): 2-4 ТБ NVMe SSD + 1 ТБ NVMe SSD

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 15/17
--	--	------------

Обоснование: SSD 1 используется для временного хранения данных Kafka и Postgres Pro. Быстрое хранилище необходимо для минимизации задержек при работе с видеопотоками.

SSD 2 используется для данных операционной системы и телеметрии.

Дополнительное долговременное хранилище (HDD или NAS): 2-4 ТБ для долговременного хранения резервных копий (не требует высоких скоростей ввода-вывода).

4.5. Логика обнаружения нарушений

Логика обнаружений нарушений представлена в виде BPMN схемы

4.6. Пользовательский интерфейс

Пользовательский интерфейс для системы видеоаналитики представляет собой веб-интерфейс, связывающийся с системой с помощью REST API. Данный интерфейс позволяет пользователю производить:

- аутентификацию;
- создание, удаление и изменение прав учётных записей пользователей;
- создание, запуск/остановка, редактирование и удаление камер;
- просмотр информации о найденных нарушениях в формате списка или таблицы.

Более подробно со способами взаимодействия с веб-интерфейсом системы можно ознакомиться на диаграмме вариантов использования (см. Рисунок 1).

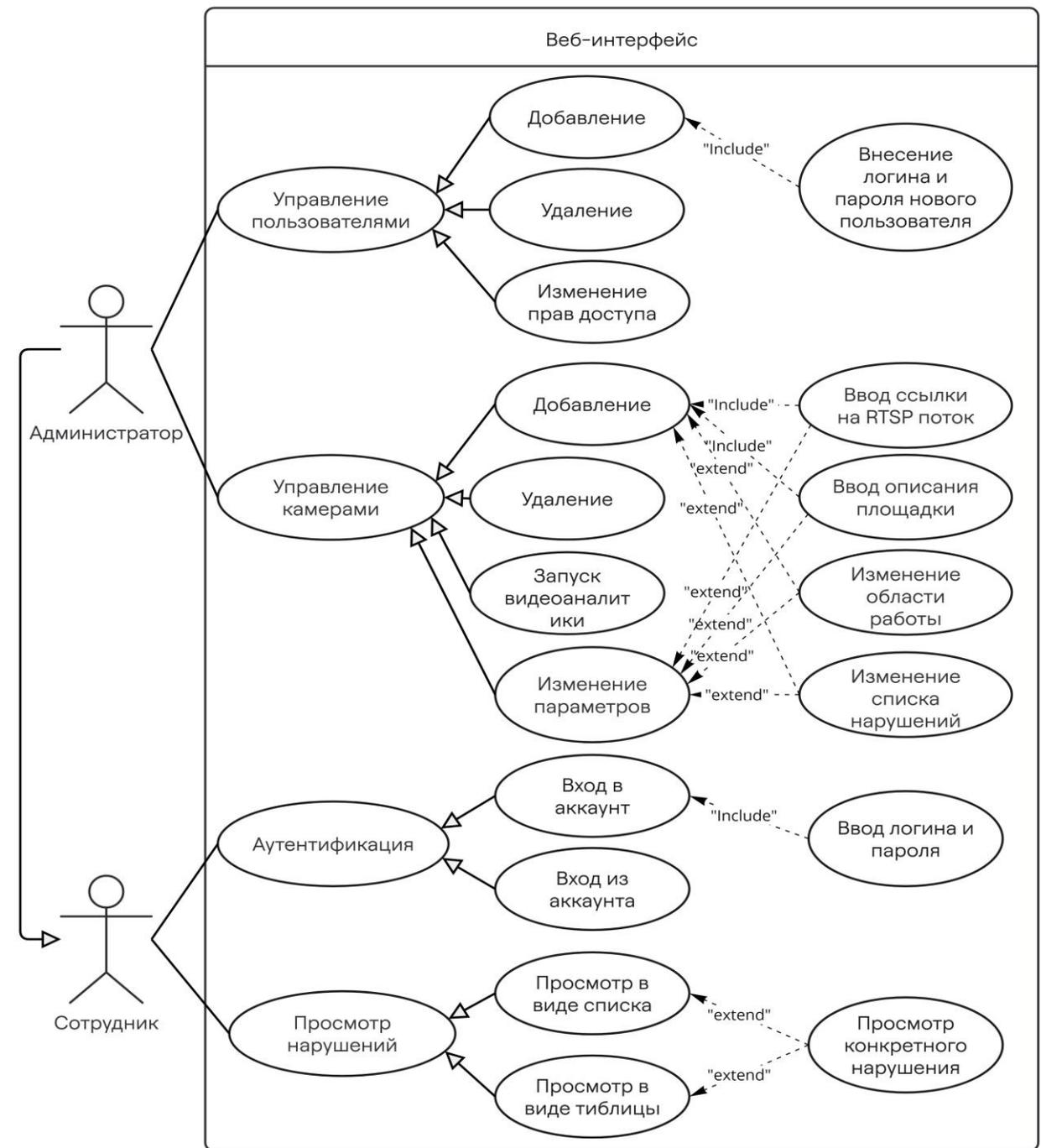


Рисунок 1 – UML диаграмма вариантов использования

Более подробно о пользовательском веб-интерфейсе можно узнать в документе «Инструкция пользователя и администратора», предоставляемого в рамках общего комплекта документации.

4.7. Безопасность

	Проектное решение. Универсальная система видеоаналитики средств индивидуальной защиты и промышленной безопасности ООО «Статанли»	Лист 17/17
--	---	------------

Взаимодействие с REST API системы производится с использованием JWT токена, предварительно выданного в рамках аутентификации. Передача данных осуществляется по протоколу HTTP в рамках закрытой сети. Использование HTTPS протокола затруднено сложностями в выдаче сертификатов. Ввиду политик безопасности закрытой сети, протокол HTTP был принят достаточным для обеспечения необходимого уровня безопасности.

Другие межкомпонентное взаимодействие производится в рамках локальной сети Docker, а сами компоненты не имеют выход во внешнюю сеть, что исключает возможность доступа к ним без получения прямого доступа к серверу.

Взаимодействие с системой InSight осуществляется с помощью HTTP протокола из за ограничений самой системы InSight, существовавших на момент реализации системы.

4.8. Лицензирование

Используемые программные компоненты системы (см. Таблица 3 не требуют лицензии или имеют лицензии с открытым исходным кодом, что позволяет их использовать без дополнительной покупки лицензий