

NannyNet

1 октября 2019 г.

Кускова Алиса

11 "Л" класс
Школа № 2026, корпус 12
Москва

Обзор проблемы (Слайд 2)

В наши дни главным приоритетом для общества становится здоровье и безопасность каждого человека.

Однако сейчас организация безопасности работает по реактивным методам: если человеку угрожает опасность, он должен позвонить в полицию и дожидаться её приезда, найти охранника, уговорить окружающих рисковать собой для помощи ему.

К сожалению, такая система безопасности работает хорошо, только когда опасное событие уже произошло (- Меня убивают! - Вот когда убьют, тогда и звоните!).

Гораздо эффективнее работают проактивные методы обеспечения безопасности, когда не допускается возникновение опасных инцидентов или вмешательство осуществляется на ранних стадиях события.

К проактивным методам, например, относятся требования техники безопасности - нельзя ходить без каски по стройке, бегать по школе, работать на станке с распущенными волосами, драться и т.д. Можно написать инструкцию, но трудно проконтролировать каждого человека и добиться их выполнения. Также невозможно

прописать все случаи жизни, поэтому в инструкциях вынуждены перестраховываться и устанавливать не всегда соответствующие здравому смыслу требования.

Избранные люди могут позволить более эффективные проактивные методы защиты: личная охрана, персональные врачи, предварительная проверка посещаемых помещений, контактов с людьми и т.д.

Но к каждому человеку охрану не поставить, поэтому в последние десятилетия стали использовать технические средства, помогающие отслеживать ситуацию.

Самым популярным средством сейчас является установка видеокамер наблюдения. Это позволяет людям, ответственным за безопасность, просматривать большую территорию и оперативно пресекать опасное поведение.

Однако на практике выяснилось, что людям сложно наблюдать за большим количеством камер (больше 4-5) длительное время (больше 20 мин), они быстро устают и теряют внимание.

В помощь были разработаны программные и аппаратные средства, которые в настоящее время крайне ограничены в своём функционале - все сводится к срабатыванию датчика или выявлению движения в кадре. Такие средства позволяют помочь в выявлении проникновения на охраняемую территорию, но никак не помогают с обеспечением безопасности в оживленных местах - в школе, детском саду, на улице, в офисе. Широкоиспользуемые датчики и камеры сами по себе не могут отличить драку от чтения книги в библиотеке, не могут отличить коньки от оружия и не предупредят, когда назревает ссора.

Цель и задачи (Слайд 3)

Цель:

Повысить надежность имеющихся систем безопасности, основанных на видеонаблюдении, и снизить нагрузку на операторов, путем разработки прототипа системы, которая способна обрабатывать данные с имеющихся систем аудио- и видеонаблюдения, анализировать их с использованием систем искусственного интеллекта, интегрировать данные, полученные различными моделями анализа, и уведомлять службу безопасности о высокой вероятности возникновения опасной ситуации, предоставляя при этом данные, позволяющие человеку перепроверить выводы системы.

Задачи:

I этап разработки (текущий проект):

1. **Разработать архитектуру решения:** спроектировать систему классов, обеспечивающих возможность поуровневой обработки и анализа входящих сигналов различными моделями, принятия решений на основе интеграции нейронных и алгоритмических моделей.
2. **Реализовать скелет решения:** реализовать конвейер, осуществляющий запуск моделей, передачу и хранение промежуточных результатов обработки.
3. **Разработать модули:** захват видеоизображения, покадровая разбивка, классификация типов объектов в кадрах и их координат, оценка поз людей, выявление опасных поз людей в статике (замах, удар, падение), выявление опасных типов объектов (оружие), оповещение при выявлении опасных ситуаций.
4. **Оценить результаты I этапа разработки:** оценить достоверность выявления потенциальной опасности происходящего на тестовых видео, оценить

производительность системы, принять решение о целесообразности дальнейшего развития проекта.

Задачи дальнейшего развития проекта:

II этап разработки:

5. **Разработать модули:** выявление типовых шаблонов поведения человека по изменению позы в динамике (бег, драка и т.д), оповещение об опасном поведении Человека (падение, драка, погоня, угрожающие жесты).

III этап разработки:

6. **Разработать модули:** определение эмоциональной окраски звуковых фрагментов, определение относительной громкости речи, оповещение об угрожающем характере речи или человеку в беде.

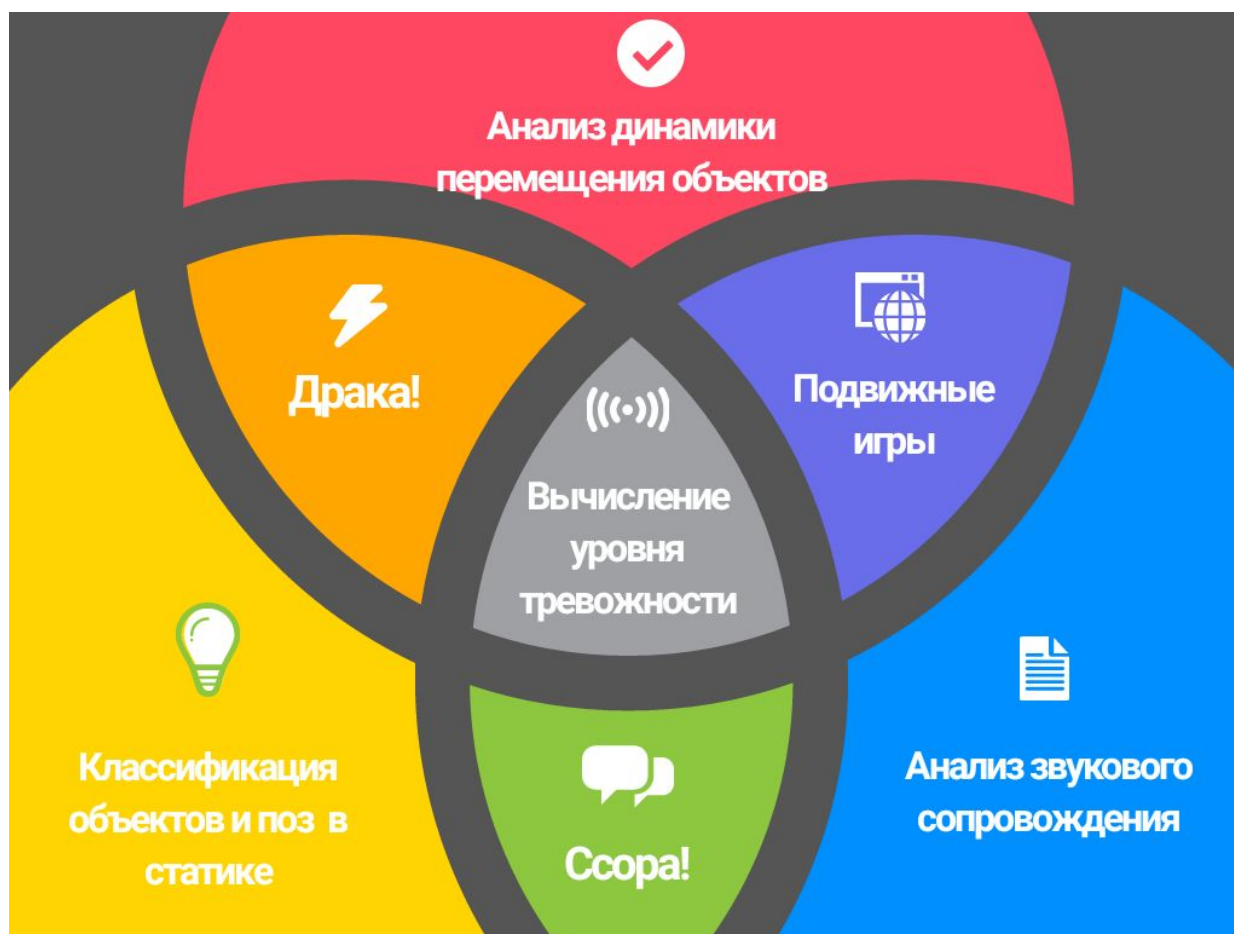
Принцип работы системы (Слайд 4)

Авторы предлагают разработать новую интеллектуальную систему проактивной безопасности в общественных местах NannyNet, использующую последние достижения в области искусственного интеллекта для выявления по существующим системам видеонаблюдения в общественных местах потенциально опасных ситуаций и информировании о них.

Такая задача является сложной даже для человека, тем более для машины.

Поэтому задачу предлагается решать построением многоуровневой модели, где на каждом уровне оцениваются различные признаки изображений и аудио с учётом результатов анализа других уровней.

Одни уровни обрабатываются при помощи традиционных алгоритмических методов, а для обработки других используются нейронные сети (для классификации объектов, оценки поз людей, эмоциональной окраски содержимого аудиопотока).



Система получает на вход видеопоток с камеры и аудиопоток с микрофона камеры.

- 1) На первом уровне находятся нейронные сети, которые получают на вход отдельные кадры с видео и небольшие куски аудио.

На этом уровне находится слой статического определения изображений (СТОЙ).

Одна нейронная сеть слоя СТОЙ выделяет на изображении отдельные объекты, классифицирует их и помечает каждый объект его типом (человек, машина, окно, птица, ружье, компьютер, портфель и т.д).

Вторая нейронная сеть слоя СТОЙ определяет позу людей (относительное положение частей рук, ног, головы и тела).

Фрагменты аудио будут передаваться в анализ аудио (АУ), где нейросеть оценивает эмоциональный окрас произносимого (спокойный разговор, угрозы, ругань, страх).

- 2) На втором уровне алгоритмическими методами обрабатываются данные первого уровня, сохраняя в памяти результаты обработки, а затем сравнивая их изменение во времени.

Таким образом алгоритм изучения динамики изображений (ИДИ) может сравнивая изменения положения, рассчитать характер их изменений и кинематические характеристики. Данные нейронных сетей первого уровня дополняются новыми характеристиками: скорость и направление движения, начало движения, окончание движения, а для объекта Человек определяются изменения позы: идёт, бежит, падает, бьет, машет руками, машет ногами.

- 3) Результаты классификации, полученные на всех слоях передаются в итоге на главный анализатор **"Need Alarm Now? No or Yes?" (NANNY)**, который на основании типа общественного места (школьный класс, коридор, спортзал, пешеходная улица, дорога, офис, стройка, военный полигон), текущего времени и дня недели (рабочее время, нерабочее время, перемена, время урока) ранжирует выявленные признаки потенциальной опасности по степени тревожности и при превышении порогового значения оповещает персонал для внимательного изучения подозрительных кадров.

Например, не так страшно, если Человек.МашетРуками, но если рядом другой Человек, то это повышает потенциальную опасность, а если рядом другой Человек.МашетНогами, то можно прийти к выводу, что назрела оочень напряжённая обстановка, особенно если слышны Оскорбления в ЧеловеческойРечи.

А вот если несколько Человек.Бегаёт и Человек.МашетРуками, и Человек.МашетНогами, но тип помещения соответствует Игровой Площадке и Время суток День и ЧеловеческаяРечь несёт эмоции Радости, то тревожность умеренная.

Но все усложняется, когда на площадке выявляется Объект.Оружие. Это может быть игрушка, а может быть настоящее. Тут при первом появлении стоит предупредить, а потом на какое-то время снизить фактор опасности оружия, но поднять чувствительность к звукам выстрелов, крикам боли и эмоциям страха и угрозы.

Алгоритмическая структура конвейера системы (Слайд 5)

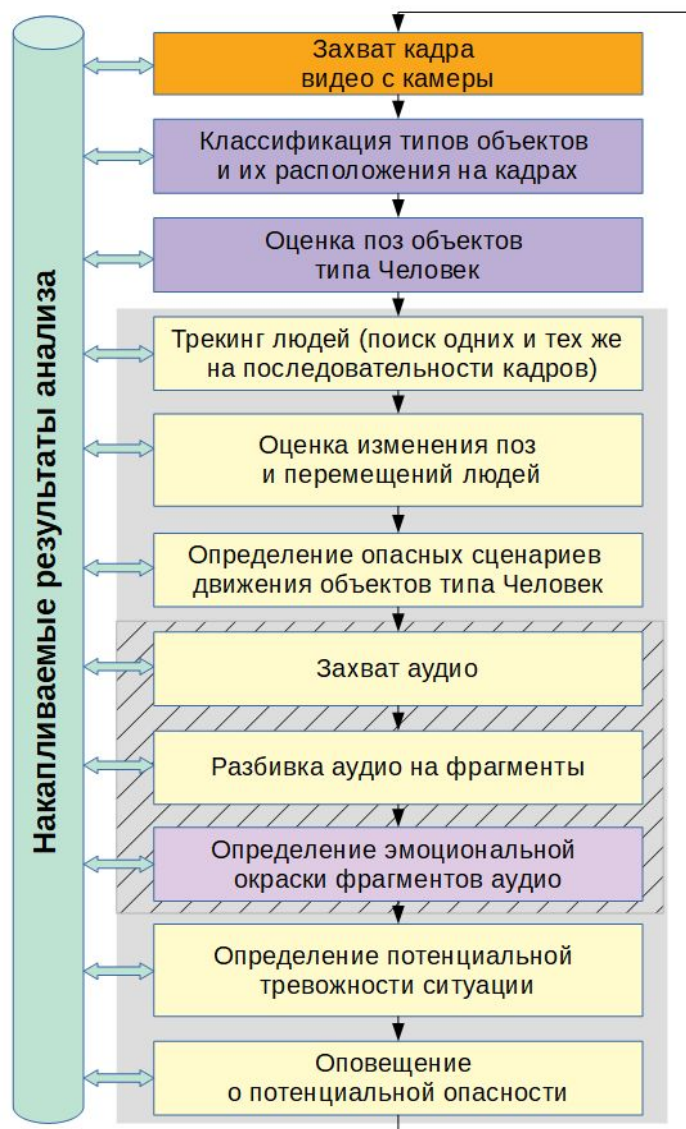
Алгоритмически система реализована в виде программных модулей, подключаемых к особому модулю - "конвейеру".

Конвейер в цикле, до тех пор, пока какой-то из модулей не сообщит о необходимости остановки, запускает по-очереди каждый модуль, обеспечивая ему доступ к результатам обработки предыдущих модулей и сохраняя результат текущего модуля в оперативный архив ограниченной длины.

Таким образом каждый модуль может получить результаты, выданные другими модулями за заданный в настройках программы период времени, а потом сделать свои выводы и записать их.

Последним модулем к конвейеру подключается главный анализатор **NANNY**, который на основании совокупности данных всех модулей принимает решение о том, чтобы сообщить людям о опасной ситуации или что оценка тревожности не выходит за допустимые границы и необходимо продолжать наблюдение за ситуацией.

Схема работы конвейера:



На текущий момент реализованы алгоритмы работы конвейера, модули обработки видео, модули реализующие работу нейронных сетей по классификации объектов в кадре и по оценке поз людей в статике.

Планируется дополнить систему модулями трекинга людей в разных кадрах, анализа изменения поз, выявления опасных движений людей, оценку степени опасности ситуации по движениям людей и наличию опасных объектов в кадре, а также оповещение оператора о возникновении опасной ситуации.

В последующих проектах система может быть дополнена модулем захвата и оценки эмоциональности аудиофрагментов.

Модульная структура системы позволяет ее легко масштабировать, распределяя нагрузку разных модулей по разным техническим средствам. Например, в школе может быть установлен компьютер с графическим процессором, который будет выполнять первичную обработку потока видео нейронными сетями, а затем вместо модулей вторичной обработки и оповещения, в конвейере будет модуль передачи обработанных данных в районные системы, где эти данные будут обработаны алгоритмическими модулями анализа ситуации.

Демонстрация работы системы (Слайд 6)

В файле `demo1_conveyor.mp4` показан вывод консоли программы, демонстрирующий последовательную работу модулей, осуществляющих покадровую обработку видео.

В файле `demo2_objects.mp4` демонстрируется графический результат работы модуля, который при помощи нейронной сети классифицирует объекты в кадре.

В файле `demo3_poses.mp4` можно увидеть как отображаются результаты распознавания поз людей на видео.

Технические и программные средства (Слайд 7)

Разработка и отладка прототипа системы осуществляется на персональном компьютере с процессором Intel Core i5, оперативная память 16 Gb, с установленной операционной системой Windows 10 64bit. На компьютере отсутствует видеоплата с графическим ускорителем, что повлияло на выбор моделей нейронных сетей для обработки изображений.

В качестве основного языка разработки программного обеспечения используется Python3, широко применяемый для подобных задач.

Разработка осуществляется в интегрированной среде Visual Studio 2019 с установленным пакетом разработки на Python. Это облегчает управление версиями кода, отладку кода и настройку библиотек среды окружения.

Обработка видео и изображений осуществляется популярной для этих целей библиотекой OpenCV.

Для работы с моделями, построенными на нейронных сетях, используется библиотека Tensorflow 1.14.

Для классификация объектов используется предобученная, свободно доступная модель, являющаяся комбинацией моделей **ssd** (отвечает за быстрый поиск объектов на изображении) и **mobilenet_v2** (отвечает за классификацию найденных объектов), обученная на изображениях из библиотеки **openimages_v4**. Эта комбинация моделей очень неприхотлива к производительности вычислительной техники и позволяет выполнять анализ в режиме реального времени.

Для оценки позы людей (относительного расположения их суставов) применена свободная библиотека **OpenPose**, использующая нейронную сеть для поиска на кадре частей тела человека и затем вычисляющая координаты суставов, связывающих их.

Для анализаторов, определяющих степень тревожности по результатам предыдущих моделей, предлагается использовать традиционные алгоритмические подходы, что позволяет более четко и однозначно задавать работу анализатора и настраивать его работу для конкретных типов помещений.

В качестве механизма оповещения предлагается использовать Telegram-бота. Такое решение позволяет легко подключать пользователей, не требует разработки отдельных приложений для пользователей, позволяет уведомлять крайне оперативно и добавлять к сообщениям кадры изображений, аудио- и видеофрагменты, которые вызвали опасения.

Заключение (Слайд 8)

Разрабатываемая система еще далека от завершения, но даже те скромные результаты анализа видео, которые удалось получить на данный момент, говорят о следующем:

NannyNet позволяет анализировать сотни предметов в кадре за десятые доли секунды, что недоступно оператору при работе с несколькими камерами. NannyNet не устает, не спит, не обедает и внимательна к каждому кадру изображения.

Система отличается от существующих систем безопасности многофакторностью своего анализа, выводы делаются на основании контекста событий - в каком типе помещения происходят события, что происходило до этого, в какое время суток это происходит, как ведут себя люди, есть ли у них при себе опасные или потенциально опасные предметы, в каком тоне происходит разговор и т.д.

В случае успешного завершения проекта, NannyNet может поднять уровень безопасности в местах, где сейчас автоматизированные средства безопасности неэффективны - в местах скопления людей - в школах, больницах, детских садах, в транспорте, в местах проведения мероприятий.

Система не может полностью заменить человека в вопросах безопасности, но она может оказать ему большую помощь, обращая внимания на детали, которые он мог пропустить и оповещая о ситуациях, которые пока не привели к беде, но могут ее спровоцировать.