

# NannyNet

1 октября 2019 г.

**Кускова Алиса**

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

## Обзор проблемы

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 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.

## Проектное предложение

Авторы предлагают разработать новую интеллектуальную систему проактивной безопасности в общественных местах NannyNet, использующую последние достижения в области искусственного интеллекта для выявления по изображениям 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 многослойной модели, где на каждом слое оцениваются различные признаки содержимого изображения с учётом результатов анализа других слоев.



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

Потоки разбиваются на небольшие куски и передаются в слои изучения динамики изображений (ИДИ) и аудио анализа (АУ), где для работы важна последовательность кадров или звуков.

Также из потока выделяются отдельные кадры, которые передаются в слой статического определения изображений (СТОЙ).

Слой СТОЙ выделяет на изображении отдельные объекты, классифицирует их и помечает каждый объект его типом.

Слоев одного типа может быть несколько, например, второй слой СТОЙ может для объектов типа Человек, выявленных на предыдущем слое, определять его позу: лежит, сидит, стоит + руки опущены, подняты + ноги стоят, подняты + голова поднята, опущена. Или для строек определять Человек в каске или Человек без каски.

Слой ИДИ обладает историей изменения расположения объектов. Сравнивая изменения положения, выявленных на предыдущем слое объектов, классифицирует эти изменения и рассчитывает их кинематические характеристики. На выходе характеристики объектов предыдущего слоя дополняются новыми характеристиками: скорость и направление движения, начало движения, окончание движения, а для объекта Человек определяются изменения позы: идёт, бежит, падает, машет руками, машет ногами.

Слой АУ анализирует аудиоканал, определяя громкости звуков, типы звуков и распознает человеческую речь. Пример результатов анализа: Глухой удар.Тихий, Разбитое стекло.Громкий, Крик.Громкий, Человеческая речь.Тихий("я тебя прирежу, понял?").

Результаты распознанных образцов речи подаются на слой лингвистической оценки восклицаний (СЛОВО), где фразам даётся эмоциональная оценка: угроза, оскорбления, положительные эмоции, нейтральная окраска.

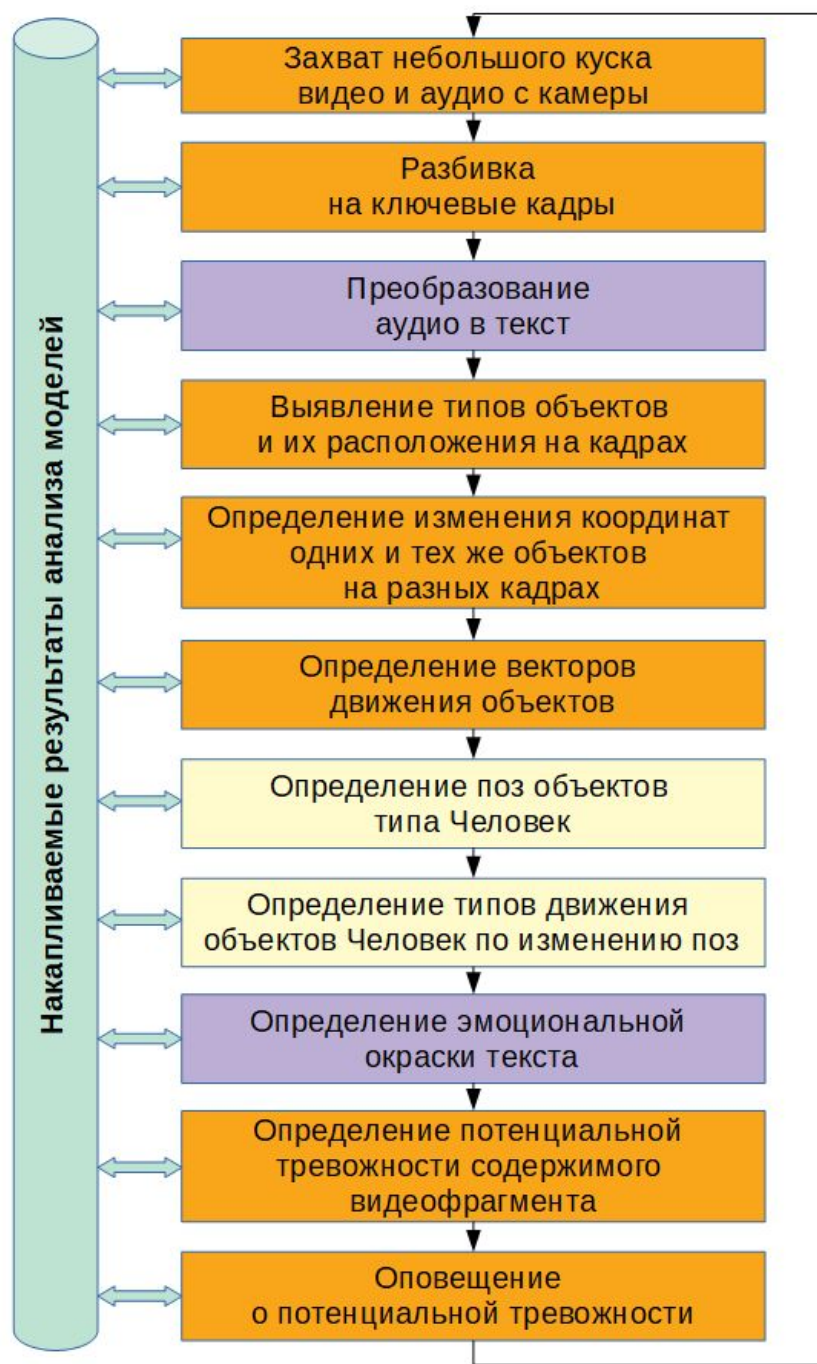
Результаты классификации, полученные на всех слоях передаются в итоге на главный анализатор "Need Alarm Now? No or Yes?" (NANNY), который на основании типа общественного места (школьный класс, коридор, спортзал, пешеходная улица, дорога, офис, стройка, военный полигон), текущего времени и дня недели (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 очень напряжённая обстановка, особенно если ЧеловеческаяРечь.Оскорбления.

А если несколько Человек.Бежит, Человек.МашетРуками, Человек.МашетНогами, но тип помещения Игровая Площадка и Время.День и ЧеловеческаяРечь.Радость, то тревожности 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.

Такая система не решит всех проблем с безопасностью, но позволит сконцентрировать внимание персонала на особо важных событиях, попавших в кадры видеокамер.



Конвейер послойной обработки

## Цели

### I этап разработки:

1. **Разработать архитектуру решения:** спроектировать систему классов, обеспечивающих возможность послойной обработки и анализа входящих сигналов различными моделями, принятия решений на основе многослойной модели.
2. **Реализовать скелет решения:** реализовать конвейер, осуществляющий запуск моделей послойно и передачу результатов обработки из внешних слоев на более глубокие.
3. **Разработать анализаторы I этапа:** захват видеоизображения, покадровая разбивка, выявления типов объектов в кадрах и их координаты, трассировка изменения координат одних и тех же объектов, вычисление скоростей движения объектов, оповещение в случае выявления в видеофрагменте опасных типов объектов (оружие) и быстродвижущихся объектов типа Человек.
4. **Проверить результаты I этапа разработки:** оценить качество определения потенциальной опасности происходящего на видео, оценить скорость работы анализаторов, определить необходимую производительность оборудования для реализации продукта применимого на практике, принять решение о целесообразности дальнейшего развития проекта на II и III этапах разработки.

### II этап разработки:

5. **Разработать анализаторы II этапа:** определение поз объектов типа Человек (расположение рук, ног, туловища относительно друг друга и туловища относительно вертикали), определение близкостоящих объектов типа Человек, выявление типовых шаблонов поведения Человека по изменению позы (бег,

удары, взмахи и т.д), оповещение об опасном поведении Человека (падение, драка, погоня, угрожающие жесты).

### III этап разработки:

6. **Разработать анализаторы III этапа:** преобразование речи на аудио в текст, определение эмоциональной окраски текста, определение относительной громкости речи, оповещение об угрожающем характере речи или человеке в беде.

## Технические подробности

В качестве основного языка разработки предлагается Python3, широко применяемых для подобных задач, включая библиотеки OpenCV для работы с видео и изображениями, Tensorflow для работы с моделями, построенными на нейронных сетях.

Для анализаторов, определяющих типы объектов на изображении, их положение, преобразования звуков в текст, предлагается использовать доступные предобученные нейронные сети, которые успешно справляются с такими нечеткими задачами.

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.