

Встроенная видеоаналитика для детектирования и сопровождения объектов при помощи многомасштабных признаков

Птицын Н.В., ООО «Синезис», pr@synesis.ru

Аннотация

Предложен новый алгоритм встроенной видеоаналитики для автоматического детектирования и сопровождения подвижных объектов в системах охранного наблюдения.

Для детектирования объектов заимствованы принципы анализа изображения из зрительной коры мозга. Использована последовательность простых пиксельных операций: линейных (взвешенное суммирование) и нелинейных (выбор максимального значения, насыщение).

Для сопровождения объектов использован гибридный алгоритм на основе (1) анализа временного ряда детектора объектов и (2) пространственной корреляции признаков текущего кадра и статистической модели объекта.

Отличительными особенностями разработанного конвейера являются вычислительная эффективность на видеопотоке высокой четкости (HD) и возможность сопровождения слабоконтрастных перекрывающихся целей на изменчивом фоне. Алгоритм реализован и внедрен на различных платформах, в том числе на сигнальных процессорах (DSP). Оценка точности встроенных алгоритмов детектирования и сопровождения произведена по методике i-LIDS. Полностью встроенная реализация на DSP одобрена i-LIDS как система первичного обнаружения для формирования оперативных тревог и для записи событий в приложениях видеонаблюдения стерильной зоны.

Ключевые слова: *встроенная видеоаналитика, первичная зрительная кора VI, многомасштабный анализ, детектирование движения, сопровождение целей, поточное видео, i-LIDS.*

1. ВСТУПЛЕНИЕ

Автоматизация обработки поточного видео в системах охранного телевидения является актуальной научно-инженерной задачей. *Видеоаналитикой* называются программное обеспечение на базе алгоритмов машинного зрения, которое позволяет детектировать, сопровождать, классифицировать и/или идентифицировать подвижные объекты в поле зрения видеокамеры без участия оператора [1], с. 287-312. Перспективным направлением видеоаналитики является встраивание алгоритмов непосредственно в камеру или ip-устройства [2]. По сравнению с серверной реализацией, *встроенная аналитика* обрабатывает видеосигнал без искажений, вносимых цифровым или аналоговым каналом связи. Такая встроенная обработка видео на более высоком разрешении и с большей частотой кадров потенциально обеспечивает более высокую точность распознавания.

С другой стороны, известные алгоритмы видеоаналитики достаточно сложно адаптировать для встроенной реализации в камерах массового производства из-за их вычислительной сложности. Аппаратные ресурсы однокристальной платформы камеры (набор процессорных команд, тактовая частота и размер памяти) лимитированы из-за ограничений по тепловыделению и стоимости. Таким образом, необходима

разработка принципиально новых алгоритмов с большей вычислительной эффективностью, особенно, если в камере используется сенсор высокой четкости.

Главная задача встроенной аналитики – обеспечить первичное обнаружение и сопровождение целевого объекта в поле зрения камеры. Результатом работы такой аналитики являются координаты, траектория движения и признаки объектов. Другие задачи, такие как уточняющая классификация, идентификация и межкамерное сопровождение могут быть эффективно выполнены на стороне сервера.

2. КЛАССИЧЕСКИЙ ПОДХОД К ДЕТЕКТИРОВАНИЮ ОБЪЕКТОВ

Общий алгоритмический подход к детектированию подвижных объектов состоит в анализе различий между текущим кадром и моделью фона. Упрощенно, этот подход называется *вычитанием фона* (background subtraction). Принадлежность пикселя к объекту (или к фону) определяется на основе отклонения значения (яркости) пикселя в текущем кадре от его статистической оценки в модели фона. Известно множество методов моделирования фона изображения [3]. Наиболее распространены методы бегущего гауссовского среднего (running Gaussian average) и гауссовой смеси (mixture of Gaussians):

1. Метод бегущего среднего успешно работает в стерильных условиях, когда фон остается неподвижным. В случае глобального изменчивого фона, например, движения леса, кустарников и воды, одномодальная модель не позволяет эффективно обнаруживать инородные объекты.
2. Метод гауссовской смеси является многомодальным и позволяет более точно описывать статистику изменчивого фона. Однако, при моделировании отдельных пикселей, а не их паттернов, он не обеспечивает заметного прироста в точности детектирования. При этом метод гауссовской смеси является существенно более ресурсоемким, чем бегущее среднее, и обычно не вписывается в рамки вычислительных возможностей встроенного процессора.

На этапе сегментирования, отдельные пиксели, выделенные из фона, объединяются в регионы при помощи морфологических операций [4], с. 481-495. Регионы, соответствующие заданным габаритам и форме, можно считать целевыми объектам. Как правило, именно фаза морфологического анализа составляет наибольшие вычислительные затраты. Сложность анализа нелинейно возрастает с увеличением площади регионов и их количества. Основные проблемы встроенной аналитики на базе классического подхода состоят в следующем:

1. Нелинейный рост алгоритмической сложности при увеличении пиксельного размера кадра и/или целевых объектов. Так, большинство встроенных алгоритмов работают на разрешении от 160 x 120 до 320 x 240 пикселей и практически не могут быть использованы

режиме высокой четкости 1920 x 1080 пикселей. Нелинейный рост сложности обусловлен морфологическими операциями детектора при объединении крупных регионов. Это ограничение не позволяет воспользоваться потенциалом мегапиксельных камер и увеличить дальность действия видеоаналитики.

2. Недостаточная точность детектирования объектов на изменчивом фоне в условиях слабой контактности.
3. Высокая частота ложных срабатываний, обусловленных естественными проявлениями окружающей среды (облака, ветер, снег, дождь, птицы и насекомые).

3. НОВЫЙ АЛГОРИТМ

3.1 Нейробиологический механизм

Идея предлагаемого алгоритма заимствована у природы, где эволюция нервной системы живых существ достигла выдающихся результатов в области видеоанализа [5]. Рассмотрим функциональную схему первичной зрительной коры V1 (рис. 1), развитой, в частности, у приматов и человека. Нейронная сеть состоит из клеток двух типов:

1. Простые клетки, обозначаемые буквой S от англ. simple, осуществляют линейную операцию взвешенного суммирования (**sum**), то есть двумерную свертку:

$$L_{m,n}^{out} = \sum_u \sum_v w_{u,v} L_{m-u,n-v}^{in},$$

где $L_{m,n}^{out}$ – выходной сигнал в точке (m, n) , $w_{u,v}$ – весовой коэффициент (ядро свертки) в окрестной точке (u, v) , $L_{m+u,n+v}^{in}$ – входной сигнал в точке $(m + u, n + v)$.

2. Сложные клетки, обозначаемые буквой C от англ. complex, осуществляют нелинейную операцию выбора максимального значения (**max**):

$$L_{m,n}^{out} = \max_u \max_v L_{m-u,n-v}^{in}, \quad u^2 + v^2 \leq r^2.$$

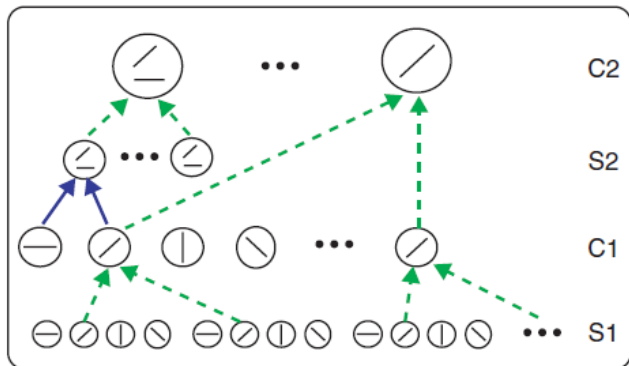


Рис 1: Обработка изображения в первичной зрительной коре: черточки – признаки, соответствующие направленным фильтрам; S1, S2 – слои простых клеток; C1, C2 – слои сложных клеток; сплошная синяя линия – взвешенное суммирование; штриховая зеленая линия – операция выбора максимума.

На входе нейронной сети исходное изображение с сетчатки глаза обрабатывается простыми клетками S1. Простые клетки реализуют направленную фильтрацию, которая выделяет границы определенных ориентаций. Задача направленного фильтра – выделить характерные признаки изображения, инвариантные к освещению. На рис. 1 показано четыре направления: горизонтальное, вертикальное и два диагональных. В области машинного зрения широко применяются аналогичные детекторы границ на базе градиента [4], с. 315-338.

На уровне сложных клеток C1 происходит группировка простых клеток S1 по каждому направлению и выделение максимального значения. Сложная клетка обладает избирательностью по признаку и обеспечивает инвариантность относительно смещения входного нейрона в окрестности группировки.

На уровне простых клеток S2 производится взвешенное суммирование выходов сложных клеток C1. В результате суммирования сигналов по различным признакам с уровня S2 получаются композитные признаки, объединяющие локальные данные по нескольким направлениям. Они похожи на признаки Хаар, но за счет предшествующего слоя сложных клеток лучше обобщают деформируемые формы.

На уровне сложных клеток C2 опять применяется нелинейная операция **max**, причем группируются не только выходы предшествующего уровня S2 (инвариантность позиции), но и выходы еще более нижнего слоя C1 (инвариантность масштаба). Таким образом, на уровне C2 простые и сложные признаки объединяются для достижения инвариантности и по смещению и по масштабу одновременно.

С другой стороны, важными особенностями клеток зрительной коры является (1) нелинейное поведение во времени и (2) нелинейная передача контраста [6]. Остановимся подробнее на особенности (2): нелинейное преобразование контраста выражается в насыщении (операция **saturate**) выходного значения признака на определенном уровне, что обеспечивает его нормирование в условиях неравномерной контрастности. Насыщение проявляется как у простых, так и сложных клеток.

Уравнение Нака-Руштона (Naka-Rushton) аппроксимирует передаточную характеристику насыщения (рис. 2):

$$L^{in}/L_{max}^{out} = L^{in}/(L^{in} + L_{sat}^{in}),$$

где L^{in} значение признака на входе клетки, L_{sat}^{in} – точка полунасыщения (на рис. $L_{sat}^{in} = 100$), L_{max}^{out} – значение на выходе клетки, L_{max}^{out} – максимальное выходное значение. Передаточную характеристику можно так же рассматривать как функцию активации нейрона.

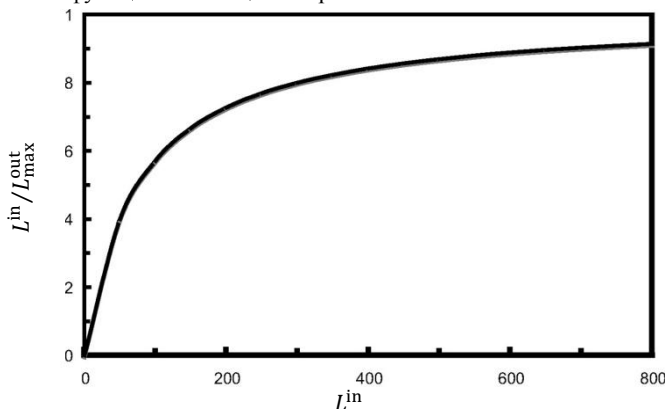


Рис 2: Передаточная характеристика клетки зрительной коры. Отметим следующие особенности зрительной коры V1 для их последующей адаптации к машинной видеоаналитике:

1. Нейронная сеть обобщает данные и понижает их размерность за счет устранения вариаций по положению и масштабу. При распространении данных от сетчатки глаза по зрительной коре понижается пространственная детализация исходного изображения, но увеличивается размерность признаков.

- Первичным признаком является не абсолютная яркость в каждой точке пространства, зарегистрированная светочувствительной клеткой, а производные признаки, полученные применением направленных фильтров к изображению.
- Простые и сложные клетки реализуют нормирование выходного сигнала за счет насыщения значения признака на некотором постоянном уровне (операция **saturate**).
- Композитные признаки получаются при помощи линейной операции взвешенного суммирования **sum** по различным признакам. Так же имеет место гауссовское суммирование сигналов, соответствующих одинаковым признакам для их обобщения.
- Инвариантность к геометрическим преобразованиям сдвига и масштабирования достигается за счет нелинейной операции **max**. При этом горизонтальное агрегирование выходов клеток одного слоя обеспечивает инвариантность относительно сдвига, а вертикальное агрегирование выходов клеток с одного или нескольких предшествующих слоев – инвариантность относительно масштаба.
- Операции **max** и **sum** применяются поочередно и итерационно. Композиция множества двух простых операций (линейной и нелинейной) порождает высокую сложность системы видеонализа в целом. Похожий прием используется в блочных алгоритмах шифрования для достижения максимальной диффузии данных внутри блока за минимальное число арифметических операций.

3.2 Многомасштабное представление

Многомасштабный подход [4], с. 125-142, уже успешно применялся для детектирования движения [7] и сегментирования сложных сцен [8]. Однако, указанные алгоритмы сегодня не пригодны для массового применения в камерах видеонаблюдения из-за их вычислительной сложности. В данной работе мы рассматриваем подходы, которые позволяют снизить ресурсоемкость алгоритмов на несколько порядков и применить алгоритмы в системах встроенного видеонализа.

Рассмотрим многомасштабное представление одного признака (простого или композитного) в виде пирамиды на рис. 3. Таких пирамид может быть несколько для каждого признака, а так же для маски сегментирования и других вспомогательных данных.

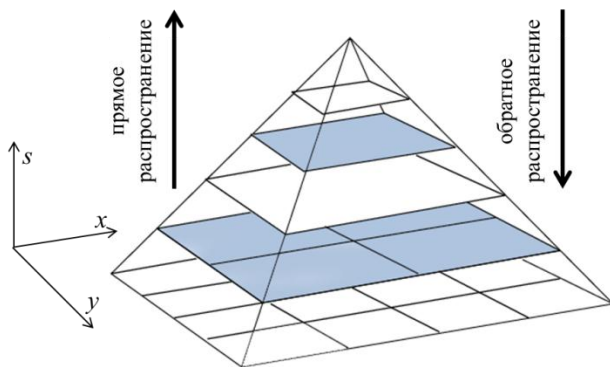


Рис 3: Две фазы многомасштабного видеонализа: x, y – пространственные координаты изображения, s – пространство масштаба.

Оптимальный алгоритм встроенной видеоаналитики включает две фазы обработки многомасштабных данных:

- Прямое распространение от детального к грубому представлению (обобщение).
- Обратное распространение от грубого к детальному представлению (уточнение). В обратной фазе обработку можно локализовать только областью детектирования и сопровождения.

В ходе прямого распространения каждый последующий слой пирамиды вычисляется при помощи простых линейных и нелинейных операций, рассмотренных выше. Операция **sum** может выступать аналогом операции **reduce** при построении гауссовой пирамиды [4], с. 137. Операция **max** усиливает характерные признаки и не дает им «размываться» в пирамиде, а так же, как обсуждалось выше, обеспечивает инвариантность относительно сдвига и масштаба. Операция **saturate** важна для устойчивой работы детекторы в условиях неравномерной освещенности и шума.

Оптимальный набор признаков и последовательность применения операций **sum**, **max** и **saturate** зависит от конкретных функций, выполняемой встроенной аналитикой:

- Для детектирования и сопровождения объектов необходим относительно небольшой набор признаков (1-4 признака). Обычно не требуются композитные признаки. С другой стороны, увеличение числа признаков, позволяет упростить алгоритм статистического моделирование фона и повысить чувствительность детектора на изменчивом фоне.
- Для распознавания типов объектов и идентификации важна хорошая репрезентативность признаков. На практике, видеоаналитический алгоритм может иметь 8-64 композитных признака.

На рис. 2 представлен результат применения машинного видеодетектора с использованием трех признаков (яркость и два насыщенных градиента). Алгоритм уверенно регистрирует уток на изменчивом фоне (водная рябь с контрастными отражениями). Используются одномодальная вероятностная модель фона, многомасштабный сегментатор, описанный ниже, и примитивный алгоритм сопровождения (связывание регионов в траекторию без построения статистической модели объекта).

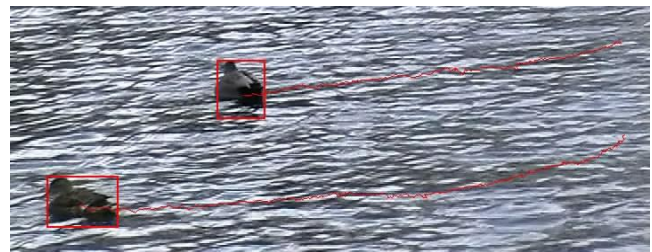


Рис 4: Сопровождение уток на изменчивом и контрастном фоне. Полное видео можно посмотреть по адресу <http://www.youtube.com/watch?v=PmJTnCIUjYw>

3.3 Сегментирование объекта

Маска объекта, рассчитываемая сегментатором, полезна для вычисления признаков объекта, для его сопровождения и для более точного моделирования фона.

Как было отмечено выше, существенным недостатком классического подхода является высокая ресурсоемкость морфологических операций на этапе определения маски регионов. Многомасштабный подход позволяет существенно повысить вычислительную эффективность видеоаналитики за счет ограничения глубины сегментирования и/или использования модели формы объекта. При этом габариты

детектируемых объектов, которые могут существенно варьироваться в поле зрения камеры, не будут значительно влиять на вычислительные затраты алгоритма.

Рассмотрим подробнее предлагаемый алгоритмический подход. В ходе *обратного распространения* сигнала в пирамиде (рис. 3), формируется многомасштабная маска региона. Сегментирование производится от грубой к детальной маске, причем процесс можно прервать после достижения желаемого уровня детализации или после исчерпания квоты вычислительных ресурсов. Входными данными являются маска с предшествующего региона, разность признаков текущего кадра и модели фона на текущем слое, а так же, опционально, модель детектируемого объекта. Выходными данными является маска региона на текущем слое. На каждом пикселе производится уточнение маски на основе метода *связывания в пирамиде* [4], с. 433-436.

На рис. 5 представлен результат работы многомасштабного сегментатора на базе метода связывания в пирамиде по нескольким признакам: сверху - исходный кадр, обрабатываемый видеоаналитикой с траекторией сопровождаемого человека; ниже - четыре слоя маски сегментирования. Маски содержат незначительные ошибки в виде точек и неточных границ, обусловленные вычислительной оптимизацией морфологических операторов.

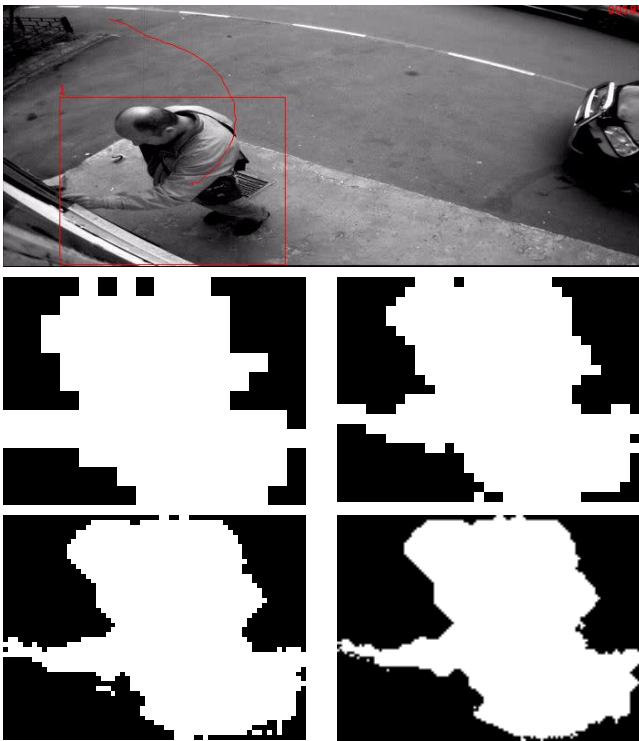


Рис 5: Многомасштабное сегментирование: исходное изображение и маски с увеличивающейся детализацией

3.4 Гибридная система сопровождения

Алгоритмы сопровождения (трассирования) позволяют получить траекторию объектов для более точного распознавания и оценки динамических признаков.

В классических реализациях встроенной аналитики популярны следующие алгоритмические подходы:

1. *Объединение регионов*, определенных детектором, во времени по множеству последовательных кадров

позволяет рассчитать искомую траекторию движения объекта. Это самый простой метод с точки зрения реализации. Основной его недостаток состоит в том, что ошибки детектора регионов приводят к нежелательным разрывам траектории. Так же, метод не позволяет сопровождать объекты в группе и сбивается при появлении ложных регионов. При медленном движении или остановке, объекты «врастают» в фон, и алгоритм сопровождения теряет цель.

2. *Корреляционные методы* предполагает построение статистической модели не только фона, но и объекта. Расчет степени схожести объекта в различных точках окрестности объекта позволяет определить наиболее вероятное его положение [4], с. 407. Преимуществом такого подхода по сравнению с первым является возможность сопровождения частично перекрывающихся объектов в группе, а так же более устойчивая работа со слабоконтрастными или медленными объектами. Основной недостаток – существенно более высокая ресурсоёмкость. Корреляционные методы неэффективны при низкой частоте кадров и сильной изменчивости сопровождаемых объектов.
3. *Оптический поток* опирается на предположение, что освещение постоянно, а форма и текстура фона и сопровождаемого объекта не меняются [4], с. 385. Метод оптического потока с вычислительной точки более эффективен, чем корреляционный метод, но уступает ему по стабильности в условиях шума и изменчивости объекта.



Рис 6: Сопровождение изменчивого объекта при помощи алгоритма связывания регионов

Более подробно методы сопровождения объектов в видеопотоке описаны в книге [4], с. 375-412.

В настоящей работе мы предлагаем гибридный метод на основе подходов (1) и (2). С одной стороны, анализ временного ряда найденных регионов позволяет эффективно сопровождать *изолированные* объекты, в том числе, которые значительно меняют форму (рис. 6). С другой стороны, корреляционный метод позволяет сопровождать объекты в группе (рис. 7) или когда детектор не обладает достаточной чувствительностью для нахождения регионов.

Объединение результатов работы алгоритмов на базе подходов (1) и (2) производится путем выбора наиболее вероятной оценки положения объекта. Для подхода (1) основой для оценки служит контрастность региона на фоне, а для подхода (2) – значения корреляции признаков по площади объекта. Подход (1) запрещается при взаимном перекрытии сопровождаемых объектов.

Многомасштабное представление маски и признаков объекта позволяет существенно повысить вычислительную эффективность корреляционного алгоритма и увеличить радиус поиска, то есть решает проблемы апертуры [4], с. 379.

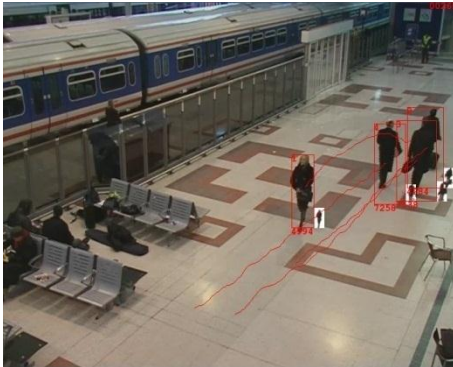
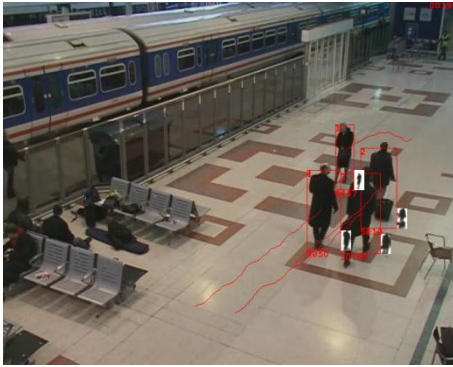


Рис 7: Индивидуальное сопровождение объектов в момент встречи (наверху) и после встречи (внизу) при помощи корреляционного алгоритма. Фрагмент видео из PETS [9].

3.5 Псевдокод конвейера видеоаналитики

Ниже представлен псевдокод конвейера разработанной видеоаналитики. Он описывает алгоритм обработки одного кадра видеопоследовательности.

Псевдокод 1 Цикл обработки одного кадра

1. Получить исходное изображение I
2. Построить гауссову пирамиду P_I из I
3. Рассчитать пирамиды градиентных и композитных признаков P_{F1}, P_{F2}, \dots из P_I (раздел 3.2)
4. Вычислить пирамиду разностей P_D между пирамидами текущего кадра P_{F1}, P_{F2}, \dots и пирамидами фона P_{B1}, P_{B2}, \dots
5. Получить маски регионов P_M из P_D с помощью алгоритма сегментирования (раздел 3.3)
6. Рассчитать траектории движения объектов гибридным алгоритмом сопровождения (раздел 3.4) из P_M (метод объединения регионов) и из P_I, P_{B1}, P_{B2} и моделей объектов (метод корреляции признаков)
7. Обновить модель фона P_{B1}, P_{B2}, \dots путем расчета гауссова среднего из P_{F1}, P_{F2}, \dots , при этом подвижные объекты маскируются с помощью P_M
8. Обновить модель объектов путем расчета гауссова среднего из P_{F1}, P_{F2}, \dots по маске P_M .

Точность и ресурсоемкость алгоритма контролируются за счет частоты кадров (допускается частичный пропуск), разрешения входного кадра, глубины сегментирования, области детектирования, области сопровождения, радиуса поиска корреляционного метода и др. настройками.

4. АППАРАТНО-ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

Видеоаналитическое обеспечение для детектирования и сопровождения объектов на базе алгоритма, приведенного в разделе 3.5, реализовано на двух аппаратных платформах: (1) x86 с использованием инструкций SSE2 для тестирования и (2) на сигнальном процессоре для встраивания непосредственно в камеру или видеозаписывающее устройство (рис. 8). Отлажено серийное производство оборудования с разработанной аналитикой.

Алгоритмы работают на всех платформах в реальном масштабе времени на разрешении от 240 линий (стандартная четкость) до 1080 линий (высокая четкость).



Рис 8: Однокристалльная реализация видеоаналитики на платформе Texas Instruments DaVinci TMS320DM6467. Размер платы, 80 x 55 x 14 мм, соответствует габаритам банковской карты.

5. МЕТОДИКА И РЕЗУЛЬТАТЫ ИСПЫТАНИЙ

Внутренние испытания видеоаналитики проводилось на специальном стенде. Разработано аппаратно-программное обеспечение для автоматического тестирования видеокамер и видеосерверов со встроенной аналитикой.

В качестве исходных материалов для тестирования использовалось видео, записанное с реальной системы уличного видеонаблюдения. Набор видеосюжетов соответствует сценарию «стерильная зона» [10] и содержит:

- 432 случая нарушения периметра (движение с различной скоростью, шагом, бегом, кувырком, ползком, в маскировочных халатах, с лестницей, в группе и по нестандартным траекториям);
- около 500 ситуаций для потенциальных ложных срабатываний (резкое изменение освещенности, движение теней, дрожание камеры, мелкие млекопитающие, птицы, насекомые на объективе, плотный снег, дождь, туман).

Набор видеосюжетов состоит из фрагментов, записанных в различное время года, суток, а так же при различных погодных условиях. Суммарная продолжительность видео – около 38 часов. Источник – стандартная камера с сенсором CCD и аналоговым выходом PAL (720 x 576 x 25 кадров/с). Цифровой формат хранения видео MJPEG, поток данных избыточный – 40 Мбит/с – что максимально приблизило качество записанного и «живого» сигналов.

Табл. 1 Точность видеоаналитики по сценарию «стерильная зона»

	Весовой параметр	Чувствительность	Специфичность	Взвешенная средняя точность
Формула	α	$r = a/(a + c)$	$p = a/(a + b)$	$F_1 = \frac{(\alpha + 1)rp}{r + \alpha p}$
Оперативная тревога	0.65	1.00	1.00	1.00
Запись событий	75.00	1.00	1.00	1.00

Разметка видеосюжетов производилась независимой группой специалистов в области безопасности и видеонаблюдения. Эксперты отмечали на видео моменты появления и исчезновения нарушителя. Аналогично были отмечены ситуации для потенциальных ложных срабатываний, что позволило эффективно классифицировать ошибки в процессе отладки. Экспертная разметка или *метаданные*, были записаны для каждого видеосюжета в формате XML для гибкого программирования текстов при помощи скриптовых языков.

Настройки алгоритмов, за исключением, калибровки глубины и области интереса, для всех видеосюжетов были идентичны. Не допускалась «подгонка» параметров алгоритмов, таких как чувствительность, под конкретные видеосюжеты.

Видеодетектор должен был зарегистрировать нарушение в течение 10 секунд. Опозданием считалось пропуском нарушения, то есть ложноотрицательным срабатыванием (счетчик c).

Повторное детектирование после разрыва траектории считалось ложноположительным срабатыванием (счетчик b). Таким образом, настоящий тест позволил оценить качество не только детектора, но и системы сопровождения.

Внутреннее тестирование показало идентичные результаты на платформе x86 и на сигнальном процессоре:

- Число истинноположительных срабатываний (нет ошибки): $a = 432$
- Число ложноположительных срабатываний (ошибка I рода): $b = 0$
- Число ложноотрицательных срабатываний (ошибка II рода): $c = 0$

Точность по сценарию «стерильная зона» рассчитана по методике i-LIDS [10]. Расчет приведен в табл. 1. Значения взвешенной средней точности для ролей оперативная тревога и записи событий совпали и оказались идеальными: $F_1=1.000$.

Видеоаналитика так же прошла внешнее независимое тестирование на другом наборе видеосюжетов, неизвестном разработчикам. Значения взвешенной средней точности для ролей оперативная тревога и записи событий составили $F_1=0.997$.

Значение точности, усредненное по результатам внутреннего и внешнего тестирования, составляет $F_1=0.999$.

6. ЗАКЛЮЧЕНИЕ

В статье представлен пример успешной адаптации нейробиологического механизма живого организма к интеллектуальным устройствам охранного телевидения. Разработаны и внедрены новые алгоритмы видеоанализа для детектирования уличного движения, сегментирования гибридного сопровождения.

Отличительной особенностью разработанных алгоритмов является использование многомасштабных признаков в виде пирамиды. Применение сразу нескольких пирамид признаков позволило одновременно отказаться от ресурсоемких алгоритмов многомодального вероятностного моделирования фона и повысить точность детектора.

Точность автоматического распознавания ситуаций на оборудовании серийного производства по сценарию «стерильная зона» составляет $F_1=1.000$ по результатам внутреннего испытания и $F_1=0.997$ по результатам независимого испытания на базе методики i-LIDS.

Перспективными направлениями для дальнейшей работы являются исследование композитивных признаков и реализация на базе них более точной классификации объектов.

7. ЛИТЕРАТУРА

- [1] Fredrik Nilsson. Intelligent network video. Understanding modern video surveillance systems, CRS Press, 2009
- [2] Птицын Н.В. *Встроенная видеоаналитика: ближайшие перспективы*, Системы безопасности, №2, 2010, с.80-83, <http://www.secuteck.ru/imag/ss-2-2010/>
- [3] Massimo Piccardi, Background subtraction techniques: a review, IEEE International Conference on Systems, Man and Cybernetics, 2004, p. 3099-3104, <http://www.utsydney.cn/www-staffit/~massimo/BackgroundSubtractionReview-Piccardi.pdf>
- [4] Bernd Jähne. Digital image processing, 5th revised and extended edition, Springer, 2002, <http://books.google.com/books?id=qUeecNvf0oC&lpg=PP1&dq=Bernd%20J%C3%A4hne.%20Digital%20image%20processing&pg=PP1#v=onepage&q&f=false>
- [5] Maximilian Riesenhuber and Tomaso Poggio, *Neural mechanisms of object recognition*, Current opinion in neurobiology, 12, 2002, p. 162–168, <http://cbcl.mit.edu/projects/cbcl/publications/ps/nb120204.pdf>
- [6] Duane G. Albrecht, Wilson S. Geisler, Robert A. Frazor and Alison M. Crane, *Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function*, Journal Neurophysiology, 88, 2002, p 888–913, <http://jn.physiology.org/cgi/content/abstract/88/2/888>
- [7] Parisa Darvish Zadeh Varcheie, Michael Sills-Lavoie and Guillaume-Alexandre Bilodeau, *A Multiscale region-based motion detection and background*, Sensors, №10, 2010, ISSN 1424-8220, <http://www.mdpi.com/1424-8220/10/2/1041/pdf>
- [8] Eitan Sharon, Meirav Galun, Dahlia Sharon, Ronen Basri and Achi Brandt, *Hierarchy and adaptivity in segmenting visual scenes*, Nature, Vol. 442, August 2006, p. 810-813,

<http://www.wisdom.weizmann.ac.il/~meirav/nature04977.pdf>

- [9] PETS: Performance evaluation of tracking and surveillance, http://www.hitech-projects.com/euprojects/cantata/datasets_cantata/dataset.html
- [10] i-LIDS User guide imagery library for intelligent detection systems, Publication №28/08 v2.0, Home Office Scientific Development Branch, p. 25-34.

Об авторе

Птицын Николай Вадимович (np@synesis.ru), выпускник МГТУ им. Н.Э. Баумана, кандидат технических наук, основатель компании «Синезис», специализирующей в области систем охранного видеонаблюдения со встроенной аналитикой.

Embedded video analytics for object detection and tracking using multiscale features

Abstract

Novel video analytics algorithms are presented enabling embedded motion detection and object tracking for CCTV systems.

The motion detection algorithm is based on the neurobiological mechanism of the primary visual cortex V1. A sequence of simple pixel operations is used including linear operators (weighted sum) and nonlinear operator (max, saturation).

The object tracking algorithm is a hybrid of two approaches: (1) time series analysis of motion detector regions and (2) space correlation between the current frame features and object model features.

Unique advantages of the present analytics pipeline include the efficiency on the high definition (HD) video stream and the ability to track low contrast overlapping objects against a dynamic background.

The embedded video analytics is implemented and deployed on different platforms including the Texas Instruments's DSP.

A comprehensive testing environment was setup used to estimate the overall performance of the video analytics implementations. A fully embedded implementation on DSP has been i-LIDS approved both as a primary detection system for operational alert use and as an event based recording system in sterile zone monitoring applications.

Keywords: *embedded video analytics, primary visual cortex V1, multiple scale features, motion detection, object tracking, i-LIDS.*

About the author

Dr Nikolai Ptitsyn (np@synesis.ru) has been graduated from Moscow State Technical University n/a N.E. Bauman and has cofounded Synesis focusing on video surveillance solutions with embedded analytics.