

АВТОМАТИЧЕСКОЕ ОБНАРУЖЕНИЕ И СОПРОВОЖДЕНИЕ ОБЪЕКТОВ В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ НА БАЗЕ ТЕХНИКИ ЧАСТИЧНОГО ДЕКОДИРОВАНИЯ H.264/AVC

Работа инициирована потребностью создания перспективных распределенных комплексов видеонаблюдения. Предложен метод автоматического обнаружения и сопровождения объектов на базе техники частичного декодирования потоков H.264/AVC.

Ключевые слова: H.264/AVC, видеоаналитика, частичное декодирование, сегментация, пространственно-временная фильтрация, марковские случайные поля, моделирование фона, MRF, векторы движения.

Постановка задачи. Задачей является реализация алгоритма обнаружения и сопровождения объектов для стационарной неподвижной камеры на базе алгоритмов частичного декодирования сигнала для гомогенной вычислительной системы – для серверного применения на базе вычислителя с единым набором инструкций (ISA). Алгоритм должен обнаруживать движущиеся объекты и распознавать их границы в видеокадре. Наблюдение происходит под произвольным углом к сцене, без прямых ограничений на изменение масштаба объектов и количество объектов, но при общем заполнении кадра объектами не более 65%. При наличии этих воздействий в тестовых материалах, необходимо сохранение показателей recall и precision [17] не ниже 70% в среднем для набора.

История вопроса. Алгоритмы видеокомпрессии используются для сжатия видео в рамках широкого спектра приложений. Выходя за рамки отлаженного процесса работы аппаратных кодеков по упаковке/визуализации видеоданных, существенным препятствием для любого рода программного анализа видеоматериала остается его размерность. Промежуточное представление информации, востребованное с точки зрения ряда алгоритмов и восстанавливаемое в результате предварительной обработки требует существенных вычислительных затрат.

Подходы к анализу кодированных данных MPEG для реконструкции движения в кадре вызвали интерес исследователей еще в конце 20 века [3, 9, 10]. Методы анализа на базе частичного декодирования (compressed domain approach, CDA-методы) на базе MPEG-2 оказались удачным решением для вычислительных машин того времени в связи с возможностью усреднения данных по макроблоку посредством использования некодированных DC-коэффициентов – поиск проводился по уменьшенному в 8 раз кадру практически без потери производительности на саму процедуру. После 2004 года большинство из этих техник стали неприменимы в условиях введения нового стандарта H.264/AVC [7].

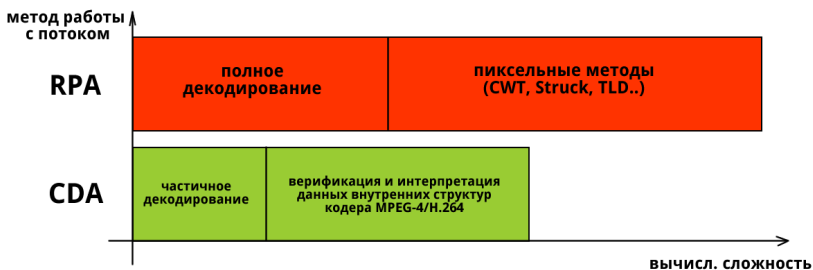


Рис. 1. Относительное распределение вычислительных ресурсов для анализа изображения на базе RPA- и CDA-методов.

В связи с этим на данный момент подавляющее большинство систем анализа движения в видеопотоке (трекеров) работают с полностью декодированными пикселями (raw pixel approach, RPA-методы). Однако, вычислительная сложность большинства RPA-методов, как например Struck[12], TLD[8], CWT[11] достаточно высока в связи с чем активно задействуются дополнительные вычислители в гетерогенной системе (GPU/DSP/..). В рамках серверной обработки это серьезно ограничивает количество обрабатываемых потоков, делает нереализуемыми задачи, связанные с жесткими требованиями по массогабаритным/энергетическим параметрам. В связи с этим предсказуемой является вторая волна CDA-методов, адаптированных для работы с H.264/AVC.

Пространственно-временной фильтр макроблоков (PSMF)

В качестве основы был выбран метод W.You, описывающий вероятностный пространственно-временной фильтр макроблоков (PSMF) (рис. 2) [15, 16].

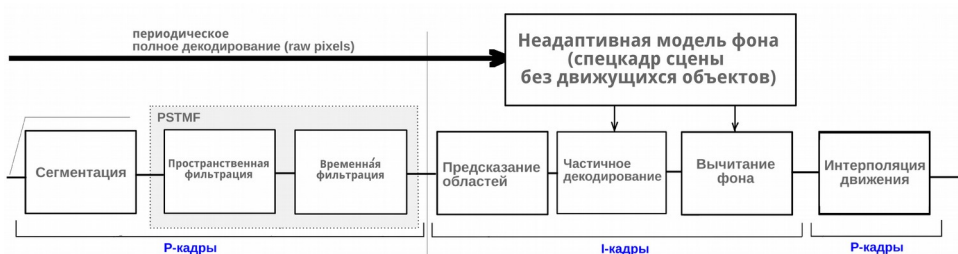


Рис. 2. Общая схема оригинального пространственно-временного фильтра макроблоков (PSMF)

Алгоритм описывает процесс фильтрации макроблоков подложки (фоновое изображение) на базе их пространственно-временных характеристик, в соответствии с которыми осуществляется сегментация регионов движения и последующий трекинг объекта. PSMF организован как кластеризация блока, пространственная и временная фильтрация.

Положительными особенностями базового алгоритма являются

- 1) автоматическое выделение движущихся объектов
- 2) инвариантность к положению камеры относительно сцены
- 3) устойчивость по отношению к структурным изменениям объекта

PSMF. Анализ макроблоков, пространственный фильтр. На этой стадии происходит основной отсев массива данных и выборка кандидатов для объектной области. В первую очередь мы убираем из списков кандидатов SKIP-макроблоки и блоки внутрикадрового кодирования 16x16 с высокой вероятностью относящиеся к фону. Остальные блоки организуются в кластеры-кандидаты, которые передаются в на следующий этап. На изображении 3 показаны различные виды блоков, отнесенных к переднему плану. Группы блоков состоят из не-SKIP блоков, объединенных по признаку пространственного соседства по горизонтали, вертикали и диагонали.

Пространственная фильтрация удаляет большинство блоков фона на основании интенсивности коэффициентов обратного преобразования Фурье (IT-коэффициенты).

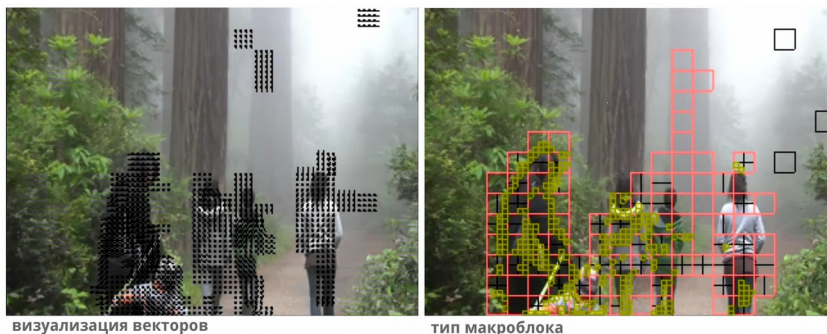


Рис. 3. Карта активных векторов (слева) и поле макроблоков (справа)

Проверка на ненулевые коэффициенты не-SKIP макроблоков позволяет:

- провести дополнительную фильтрацию шумовых флуктуаций, убрать из списка кандидатов на объекты часть блоков.
- вернуть в список кандидатов в объекты часть блоков, в случае если они были помечены как SKIP, однако содержат существенно ненулевые IT-коэффициенты (10% от максимального уровня).

PSMF. Временной анализ макроблоков. На этапе временной фильтрации мы осматриваем временную стабильность параметров каждого кандидата в объекты. В случае сохранения данных в рамках периода наблюдения кандидату присваивается статус объекта.

PSMF. Выделение объекта, коррекция траектории. На практике, существует неточность в определении объекта рамкой, в связи с чем мы обновляем позицию прямоугольника на каждом кадре в рамках определения и слежения. Для этого мы используем вычитание фона и интерполяцию

движения. Так, мы периодически обновляем данные о размерах и местоположении объекта в каждом GOP через вычитание фона. Вычитание производится на каждом I-кадре путем его сравнения с фоновым изображением, после чего производится уточнение для реального объекта. Затем выполняется интерполяция движения по P-кадрам в соответствии с данными ближайших I-кадров (см. рис. 4). После формирования модели фона, она становится активной до следующей намеченной ревизии.

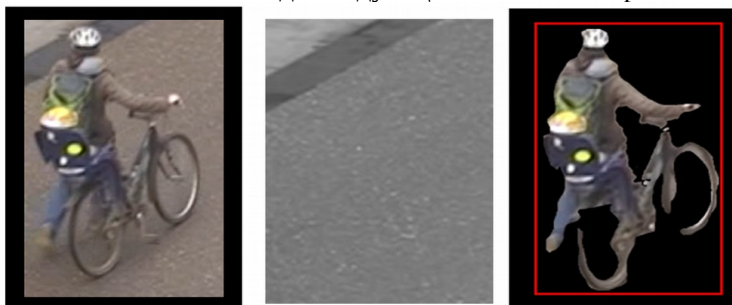


Рис. 4. Реконструкция объекта на базе модели фона

При тестировании алгоритма наблюдалось дрожание рамки объекта, что было вызвано как прошедшей через фильтр флуктуацией размеров (не была устранена на предыдущих этапах), так и неизбежным дрожанием контура, вызванным кратковременным захватом случайных бловок-кандидатов, находящихся на стадии прохождения временного фильтра.

Мофицированный пространственно-временной фильтр макроблоков (MV-PSMF)

PSMF-метод содержит в себе ряд существенных недостатков, наличие которых не позволяет использовать его в рамках современной инфраструктуры видеоанализа:

- 1) работа только с базовым профилем кодирования H.264/AVC
- 2) необходимая высокая частота ключевых кадров ($S_{GOP} < 10$)
- 3) фиксированная структура GOP (IPP...PPI)
- 4) невысокий уровень надежности обнаружения при наличии стабильных шумовых флуктуаций в кадре
- 5) расширение границ объектов за счёт флуктуаций окружения.
- 6) отсутствие динамической адаптации фонового изображения по отношению к изменениям сцены

Отмеченные недостатки приводят к существенным искажениям работы фильтра. В рамках тестовых наборов показатели PSMF не позволяли использовать его для обнаружения и сопровождения объектов (см. таблицу).

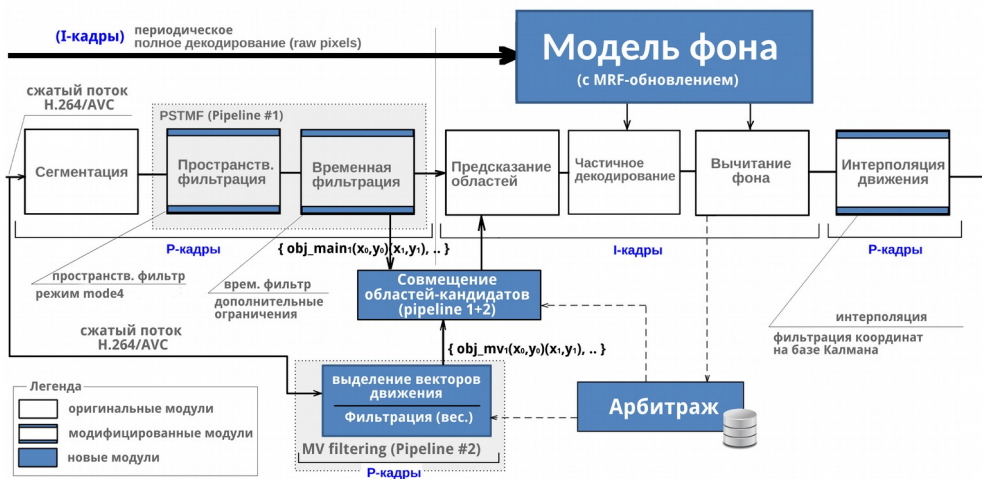


Рис. 5. Общая схема модифицированного пространственно-временного фильтра макроблоков с конвейером анализа движения (MV-PSMF)

Предлагаемый в данной статье модифицированный метод MV-PSMF (рис. 5) является следующим шагом в сторону реализации конкурентного анализатора на базе CDA-подхода.

MV-PSMF. Формирование движущихся объектов. На ряде тестовых видео PSMF выдает большое число ложных срабатываний на устойчивых флуктуациях в затемненных частях кадра, определения качающихся деревьев в качестве объектов, что приводит к срыву сопровождения реальных объектов, в случае их близости с ложными зонами. Также наблюдаются существенные неточности определения границ объекта, которые поглощая соседние макроблоки с ошибочными выбросами, существенно разрастаются, иногда до размеров кадра. Для преодоления этих недостатков вводится второй канал анализа данных на основе векторов движения. На рис. 4 (слева) изображены первичные вектора, полученные в результате отсека шумовых данных о движении.

MV-PSMF. Совмещение областей-кандидатов. На основе обработки данных с двух конвейеров (PSMF и MV), происходит формирование области объекта. Основной новой области являются область на выходе PSMF, которая корректируется при помощи данных о движении в данном секторе. Для областей с большими и однородными векторами движения выполняется операция дилатации, в результате которой образуется новая область, подтверждающая данные конвейера #1 (PSMF) (см. рис. 6).

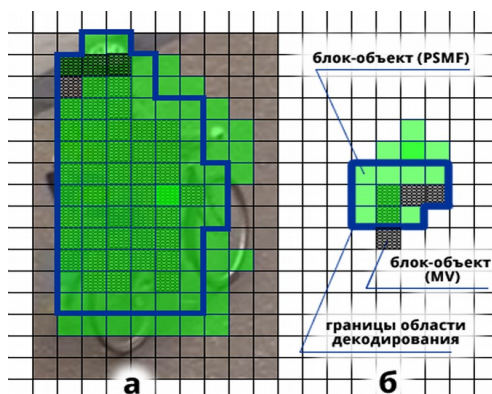


Рис. 6. Разметка границ для частичного декодирования объекта-кандидата (тестовые данные — а, пример выполнения правила — б)

MV-PSMF. Построение адаптивной модели фона. В случае видеонаблюдения вне помещения существенную роль играют эффекты окружающей среды, вызывающие блики, неоднородность воздушного пространства, перекрытия и т. д. В связи с этим наиболее актуальной задачей для корректной работы основного тракта является актуализация модели фона.

Алгоритм формирования модели фона на базе MRF [13] не требует хранения всей выборки кадров для реконструкции фона. Так как в базу данных заносятся только рабочие блоки, алгоритм обладает более высокой масштабируемостью и быстродействием, нежели альтернативные решения. Реализация на базе последовательного анализа секторов в конфигурации разбиения 2×2 с перебором блоков-кандидатов на четвертый квадрант происходит по принципу минимизации частотных возмущений (рис. 7). Замена алгоритма обеспечила более высокое качество модели фона.

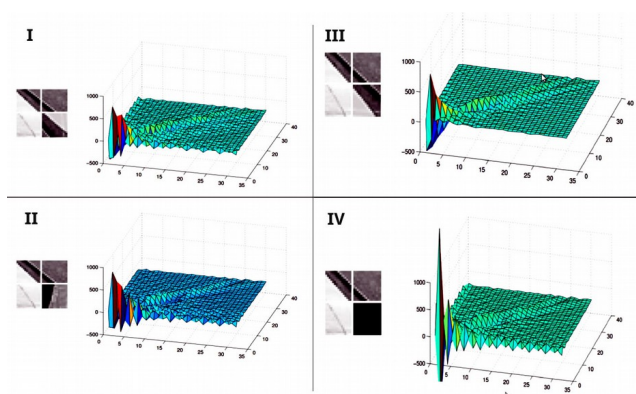


Рис. 7. Частотные возмущения для различных кандидатов [13].

Описанные в конце статьи результаты по производительности уже содержат реализацию итеративного MRF-обновления модели фона раз в 10 секунд.

MV-PSMF. Модуль арбитража. Модуль управления добавлен для осуществления контроля качества работы и согласования отдельных элементов пространственно-временного фильтра, потока и подсистем формирования фона:

1) Модуль формирования модели фона – принудительное обновление фона (при существенных изменениях в результатах сегментации, происходящих в неограниченных зонах видеокадра)

2) Хранение информации об объектах, ведение базы данных

3) Управление граничными режимами работы:

- Детектируемая площадь объекта не больше 30% кадра.
- Ограничение увеличения масштаба объекта более чем на 20% в секунду при отсутствии масштабирования (фронтальная съемка на автостраде).

Быстродействие и качество. Сравнение PSMF и MV-PSMF

В качестве основы для разработки фильтра частичного декодирования была выбрана библиотека JM-19.0 [5]. Программа формирования фона после тестирования также была встроена в приложение как программный модуль.

Для наиболее релевантного сравнения параметров производительности с результатом [16] был организован тестовый стенд с аналогичным оборудованием (Pentium 4 3.0GHz, 1 Гб RAM) где для оригинальных тестовых видео были получены результаты 49.5 кадров/сек. внутри помещения - и 37.1 вне помещения. Предложенные методы позволили увеличить эти показатели до 64.3 и 50.1 кадров/сек., соответственно.

Отслеживание качества работы алгоритма относительно PSMF рассматривалось на базе ряда численных показателей, реализованных в виде программной подсистемы автоматического определения качества сопровождения [2, 14], где вводится схема построения precision/recall-кривых [32]. Для этого каждому кадру видео с траекторией движения объекта присваивается метка DETECT, если пересечение построенного и размеченного окаймляющего прямоугольника больше некоторого порога. Кривая задается как неявная функция порогового значения:

$$Precision = \frac{|DETECT|}{|TR_{frames}|} \quad \text{и} \quad Recall = \frac{|DETECT|}{|TRAJ_{frames}|} \quad (1),$$

где $|TR_{frames}|$ – количество кадров видео, на которых построена траектория, $|TRAJ_{frames}|$ – количество кадров, на которых размечена траектория.

Тестовая выборка состояла из наборов CAVIAR [7] и IVY LAB dataset [10], по 5 последовательностей из источника. Также была исследована собственная подборка данных видеонаблюдений SURV, отличающаяся от [7] и [10] высоким разрешением (до 1920x1080) и повышенной детализацией. Сравнительный анализ алгоритмов с эталонной моделью фона производился на базе скрипта на языке Python, целью которого было сравнение XML-

файлов от каждого набора с выводом тестового частичного декодера на базе JM-19.0. Результаты работы скрипта приведены в таблице.

Таблица. Сравнение свойств оригинальной модели и MV-PSMF

	Оригинальный метод (PSMF)			Предложенный метод (MV-PSMF)		
	CAVIAR	IVY	SURV	CAVIAR	IVY	SURV
Precision	44.4%	61.3%	35.2%	72.4%	78%	52%
Recall	52.2%	65.4%	37.1%	84%	67%	56%

Выводы. Для повышения параметров качества работы в условиях стационарной камеры в алгоритм был добавлена модель актуализации фона на базе марковских случайных полей (MRF), дополнительный конвейер анализа данных движения и модуль арбитража. Эти шаги позволили обеспечить адаптивность к внешним воздействиям, за счёт чего характеристики системы были существенно улучшены, в среднем на 27% по показателю Precision и на 14% по Recall. В результате оптимизации конвейера скорость обработки возросла в среднем на 30%, что обеспечивает обработку 4-5 CIF-потоков в режиме реального времени для современного офисного компьютера (Core2Duo). Достигнутые показатели позволяют использовать алгоритм для децентрализованных систем анализа видеoinформации.

Литература

1. A. Benzougar, P. Bouthemy, and R. Fablet, *MRF-based moving object detection from MPEG coded video*. – in *Proc. IEEE Int. Conf. Image Processing, 2001*, vol. 3, pp.402-405.
2. *An open source tracking evaluation web site [Электронный ресурс]*. – Режим доступа: <http://www.cs.cmu.edu/~rcollins/Papers/opensourceweb.pdf>. (05.04.2016)
3. C.D. Creusere, G. Dahman. *Object detection and localization in compressed video*. – *Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, 2001*, vol. 1, pp. 93-97.
4. *CAVIAR Test Case Scenarios. [Электронный ресурс]*. – Режим доступа: <http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1>. (02.04.2016)
5. *H.264/AVC Software Coordination. [Электронный ресурс]*. – Режим доступа: <http://iphome.hhi.de/suehring/tml/>. (02.04.2016)
6. Hosik Sohn, Wesley De Neve, Yong Man Ro. *Privacy Protection in Video Surveillance Systems: Analysis of Subband-Adaptive Scrambling in JPEG XR*. – *Circuits and Systems for Video Technology, IEEE Transactions on*, Feb. 2011, vol.21, no.2, pp.170-177.
7. Iain E. Richardson, *H.264 and MPEG-4 Video Compression Video Coding for Next-generation Multimedia*, Wiley 1 edition, 2003.
8. Kalal, K. Mikolajczyk, and J. Matas. *Tracking-Learning-Detection*. – *Pattern Analysis and Machine Intelligence 2011*, 2011.
9. L. Favalli, A. Mecocci, F. Moschetti. *Object tracking for retrieval applications in MPEG-2*. – *IEEE Transactions on Circuits and Systems for Video Technology*, v.10 n.3., April 2000, p.427-432

10. O. Sukmarg, K.R. Rao, *Fast object detection and segmentation in MPEG compressed domain – IEEE TENCON, Kuala Lumpur, Malaysia, 2000, vol. 3, pp. 364-368.*
11. P. Brault. *A new scheme for object-oriented video compression and scene analysis based on motion-tuned spatio-temporal wavelet family and trajectory identification. – IEEE ISSPIT 2003, International Symposium on Signal Processing and Information Technology, 2003, c. 250– 254.*
12. Sam Hare, Stuart Golodetz, Amir Saffari and etc. *Struck: Structured Output Tracking with Kernels. – IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 12.2015.*
13. Vikas Reddy, Conrad Sanderson, Brian C. Lovell. *A Low-Complexity Algorithm for Static Background Estimation from Cluttered Image Sequences in Surveillance Contexts. EURASIP J. Image and Video Processing 2011, pp. 132-139.*
14. *VIVID Tracking Evaluation Web Site [Электронный ресурс]. – Режим доступа: <http://vision.cse.psu.edu/data/vividEval/main.html>. (05.04.2016)*
15. You, Wonsang and Sabirin, MS Houari and Kim, Munchurl. *Moving object tracking in H. 264/AVC bitstream. – Multimedia Content Analysis and Mining, 2007, c. 483–492.*
16. You, Wonsang and Sabirin, MS Houari and Kim, Munchurl. *Real-time detection and tracking of multiple objects with partial decoding in H. 264/AVC bitstream domain. – IS&T/SPIE Electronic Imaging, 2009.*
17. Z. Kalal, J. Matas, K. Mikolajczyk. *Online Learning of Robust Object Detectors during Unstable Tracking. – Computer Vision Workshops, IEEE 12 th International Conference on Computer Vision, ICCV'09, 2009, pp.1417–1424.*

REAL-TIME AUTOMATIC DETECTION AND TRACKING BASED ON H.264/AVC PARTIAL DECODING

The work has been initiated in the search of advanced distributed CCTV and robotic systems components solutions. Method of the automatic object detection and tracking partial decoding H.264/AVC streams in application for fixed cameras has been offered. W. You`s implementation of probabilistic spatial-temporal filter slice (PSMF) has been used. New MV-PSMF filter has been developed for new stability, robustness and scalability level approach.

Keywords: H.264/AVC, video analytics, partial decoding, segmentation, spatial filtering, temporal filtering, Markov random field, background model, MRF, moving vectors.